CSE 461: Privacy

Ben Greenstein

Jeremy Elson

TAs: Ivan and Alper

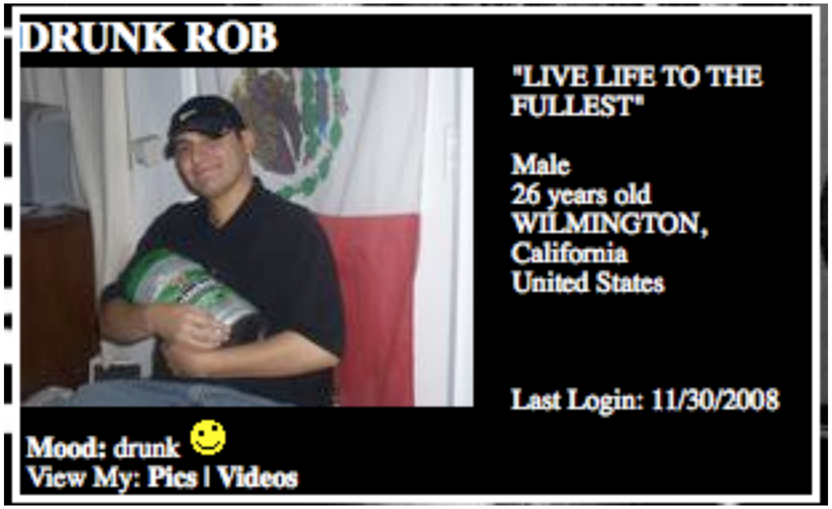
Administrivia

- Ivan has extended office hours tomorrow
- Short final review on Wednesday
- Project due this Friday
- Final exam on 12/11

- “Privacy is for old people. This is the MySpace generation. People publish every detail of their lives for all the world to see.”
  - – Mary Baker, HP

Lesson 1:

- You won’t realize privacy is important until you’re ruined by the release of personal information
  - This is dumb



DRUNK ROB

"LIVE LIFE TO THE FULLEST"

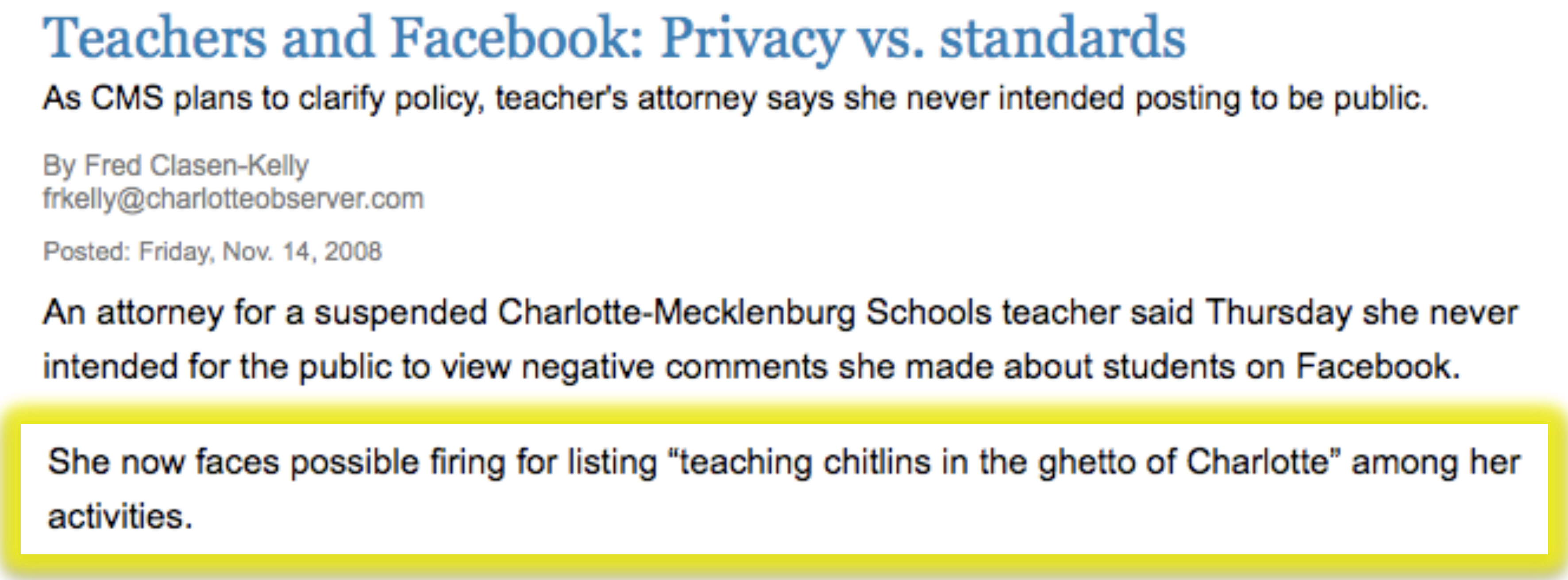
Male
26 years old
WILMINGTON,
California
United States

Last Login: 11/30/2008

Mood: drunk ☺
View My: Pics | Videos

  - Drunk Rob on MySpace

Consequence



Teachers and Facebook: Privacy vs. standards

As CMS plans to clarify policy, teacher's attorney says she never intended posting to be public.

By Fred Clasen-Kelly
frkelly@charlotteobserver.com

Posted: Friday, Nov. 14, 2008

An attorney for a suspended Charlotte-Mecklenburg Schools teacher said Thursday she never intended for the public to view negative comments she made about students on Facebook.

She now faces possible firing for listing “teaching chitlins in the ghetto of Charlotte” among her activities.

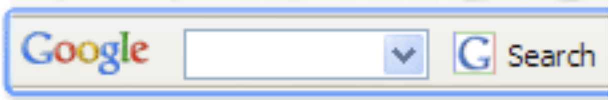
- Ok, so let’s presume you don’t deliberately put yourself at risk…

Let’s say you’re careful

- Assumptions
  - You don’t publish information about yourself
  - You use best practices of security whenever available
    - SSL/TLS-enabled Web pages
    - WPA2 with RADIUS-based authentication (enterprise)

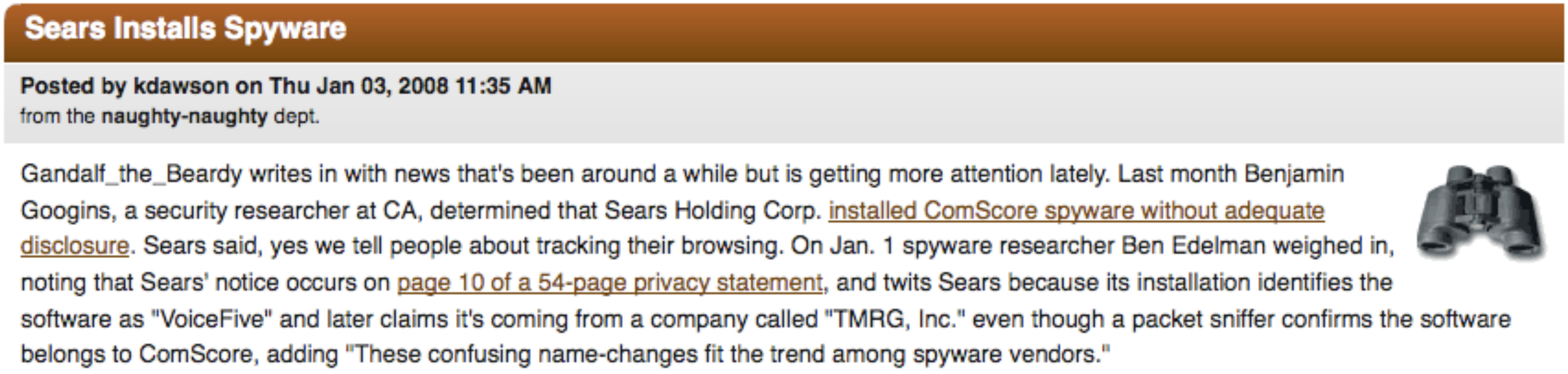
Threat 1: Big Business Web Server

- Learns who you are, where you are
- Collects all information that you tell it



Gmail™ by Google BETA

- Sometimes Maliciously

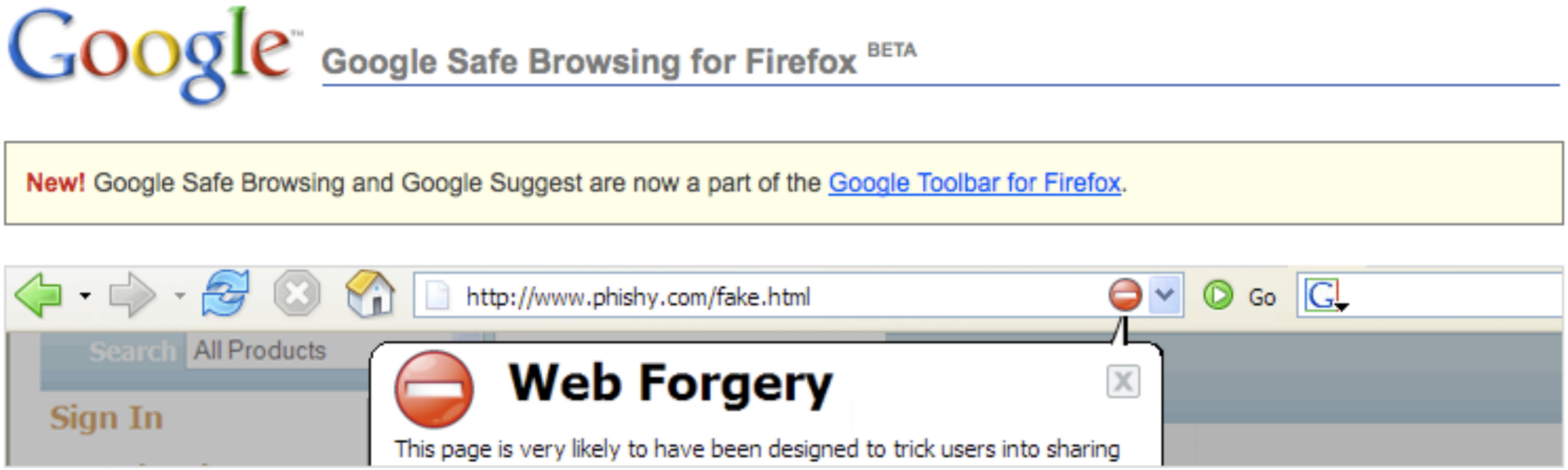


Sears Installs Spyware

Posted by kdawson on Thu Jan 03, 2008 11:35 AM
from the **naughty-naughty** dept.

Gandalf_the_Beardy writes in with news that's been around a while but is getting more attention lately. Last month Benjamin Googins, a security researcher at CA, determined that Sears Holding Corp. installed ComScore spyware without adequate disclosure. Sears said, yes we tell people about tracking their browsing. On Jan. 1 spyware researcher Ben Edelman weighed in, noting that Sears' notice occurs on page 10 of a 54-page privacy statement, and twits Sears because its installation identifies the software as "VoiceFive" and later claims it's coming from a company called "TMRG, Inc." even though a packet sniffer confirms the software belongs to ComScore, adding "These confusing name-changes fit the trend among spyware vendors."

Less obvious threat…



Google Safe Browsing for Firefox BETA

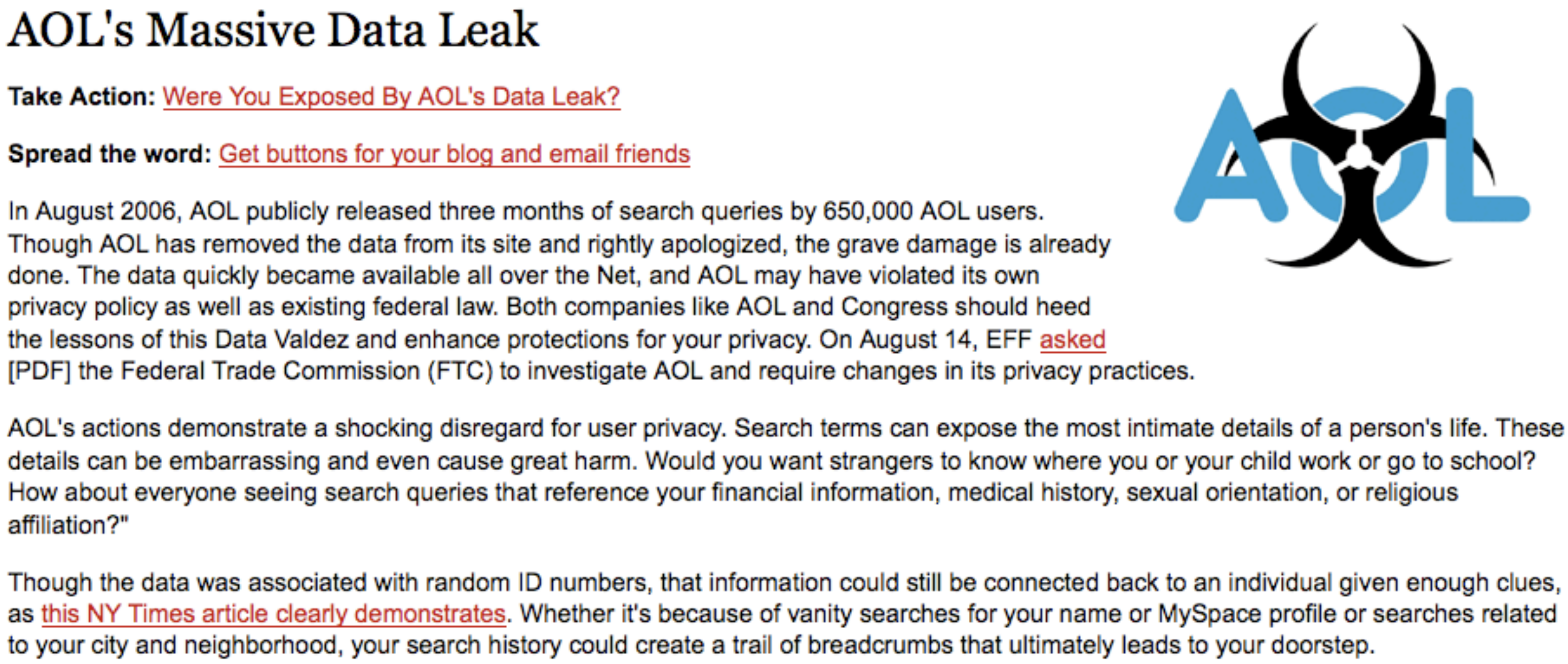
New! Google Safe Browsing and Google Suggest are now a part of the Google Toolbar for Firefox.

http://www.phishy.com/fake.html

Web Forgery

This page is very likely to have been designed to trick users into sharing

Threat doesn’t have to be intentional



AOL's Massive Data Leak

Take Action: Were You Exposed By AOL's Data Leak?

Spread the word: Get buttons for your blog and email friends

In August 2006, AOL publicly released three months of search queries by 650,000 AOL users. Though AOL has removed the data from its site and rightly apologized, the grave damage is already done. The data quickly became available all over the Net, and AOL may have violated its own privacy policy as well as existing federal law. Both companies like AOL and Congress should heed the lessons of this Data Valdez and enhance protections for your privacy. On August 14, EFF asked [PDF] the Federal Trade Commission (FTC) to investigate AOL and require changes in its privacy practices.

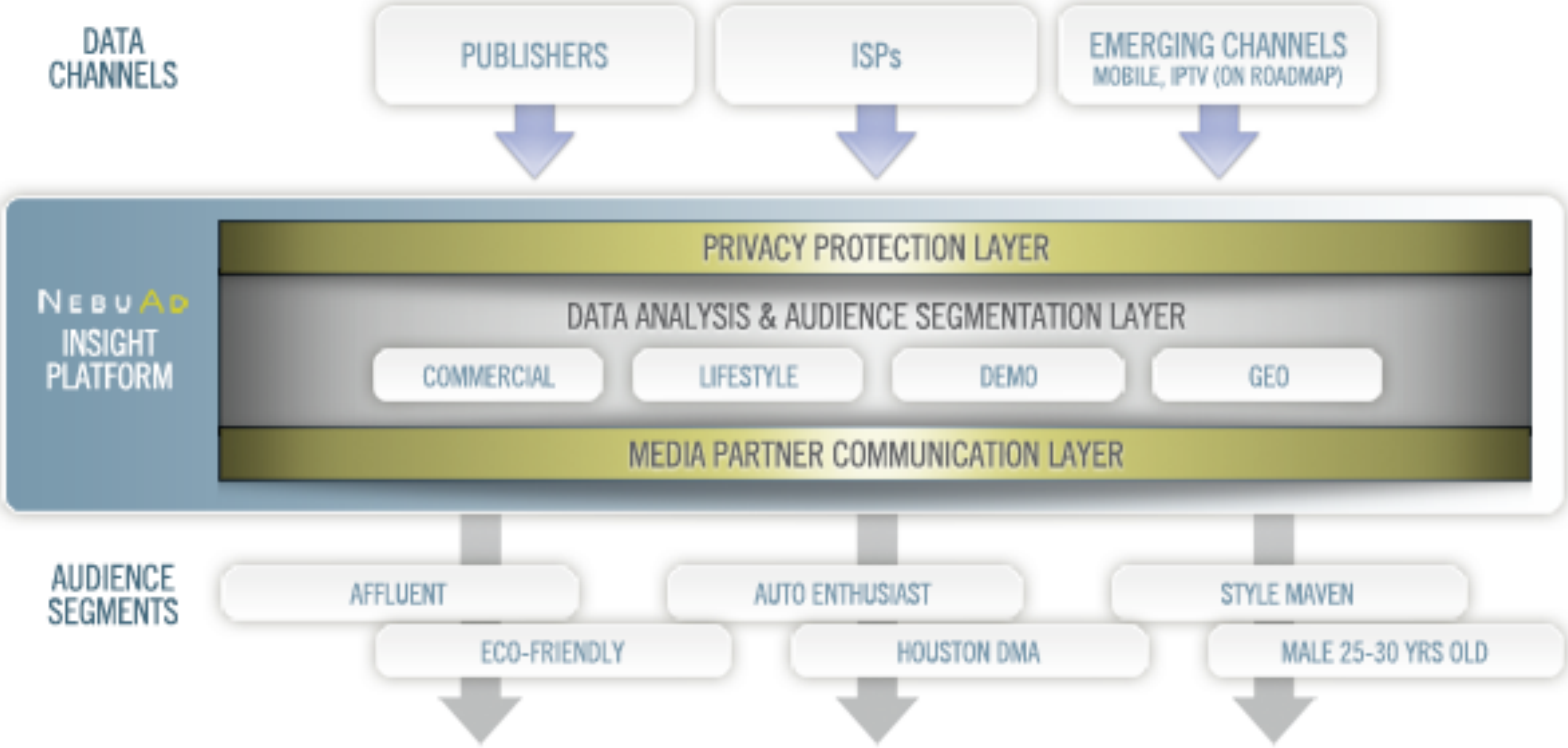
AOL's actions demonstrate a shocking disregard for user privacy. Search terms can expose the most intimate details of a person's life. These details can be embarrassing and even cause great harm. Would you want strangers to know where you or your child work or go to school? How about everyone seeing search queries that reference your financial information, medical history, sexual orientation, or religious affiliation?"

Though the data was associated with random ID numbers, that information could still be connected back to an individual given enough clues, as this NY Times article clearly demonstrates. Whether it's because of vanity searches for your name or MySpace profile or searches related to your city and neighborhood, your search history could create a trail of breadcrumbs that ultimately leads to your doorstep.

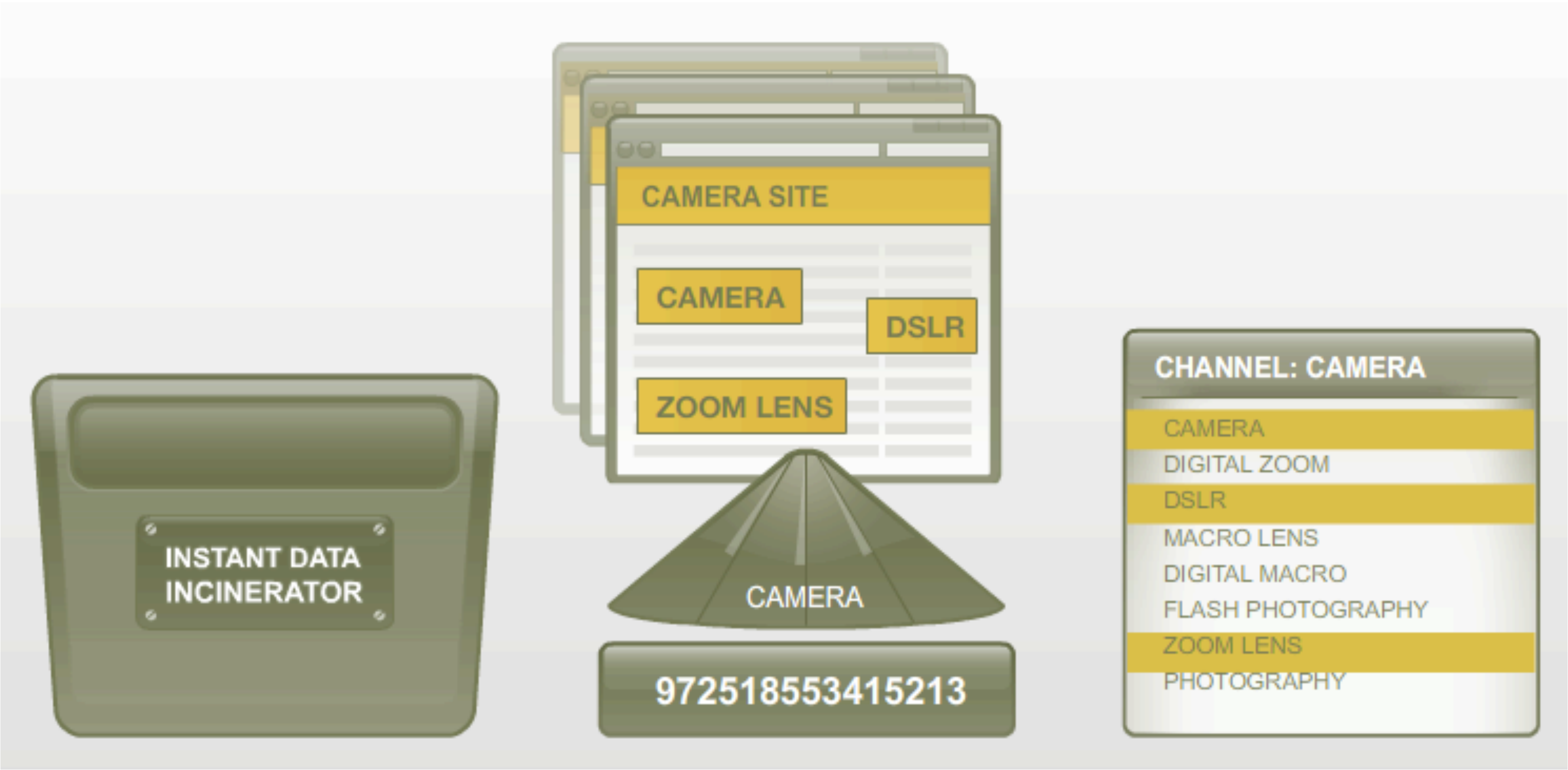
Threat 2: Mercenary ISP

- ISPs track customer browsing habits
  - Sell information to advertisers
  - Embed targeted ads in web pages (1.3%)
    - [Web Tripwires: Reis et al., 2008]
    - Example: MetroFi (free wireless)
- Technologies used for tracking at ISP
  - NebuAd, Phorm, Front Porch
  - Bring together advertisers, publishers, and ISPs
    - At ISP: inject targeted ads into non-SSL pages
- Tracking technologies at enterprise networks:
  - Vontu (semantec), Tablus (RSA), Vericept

E.g., NebuAd's Platform



E.g., Phorm


CAMERA SITE
CAMERA
DSLR
ZOOM LENS
INSTANT DATA INCINERATOR
CAMERA
972518553415213
CHANNEL: CAMERA
CAMERA
DIGITAL ZOOM
DSLR
MACRO LENS
DIGITAL MACRO
FLASH PHOTOGRAPHY
ZOOM LENS
PHOTOGRAPHY

Threat 3: Overeager Governments

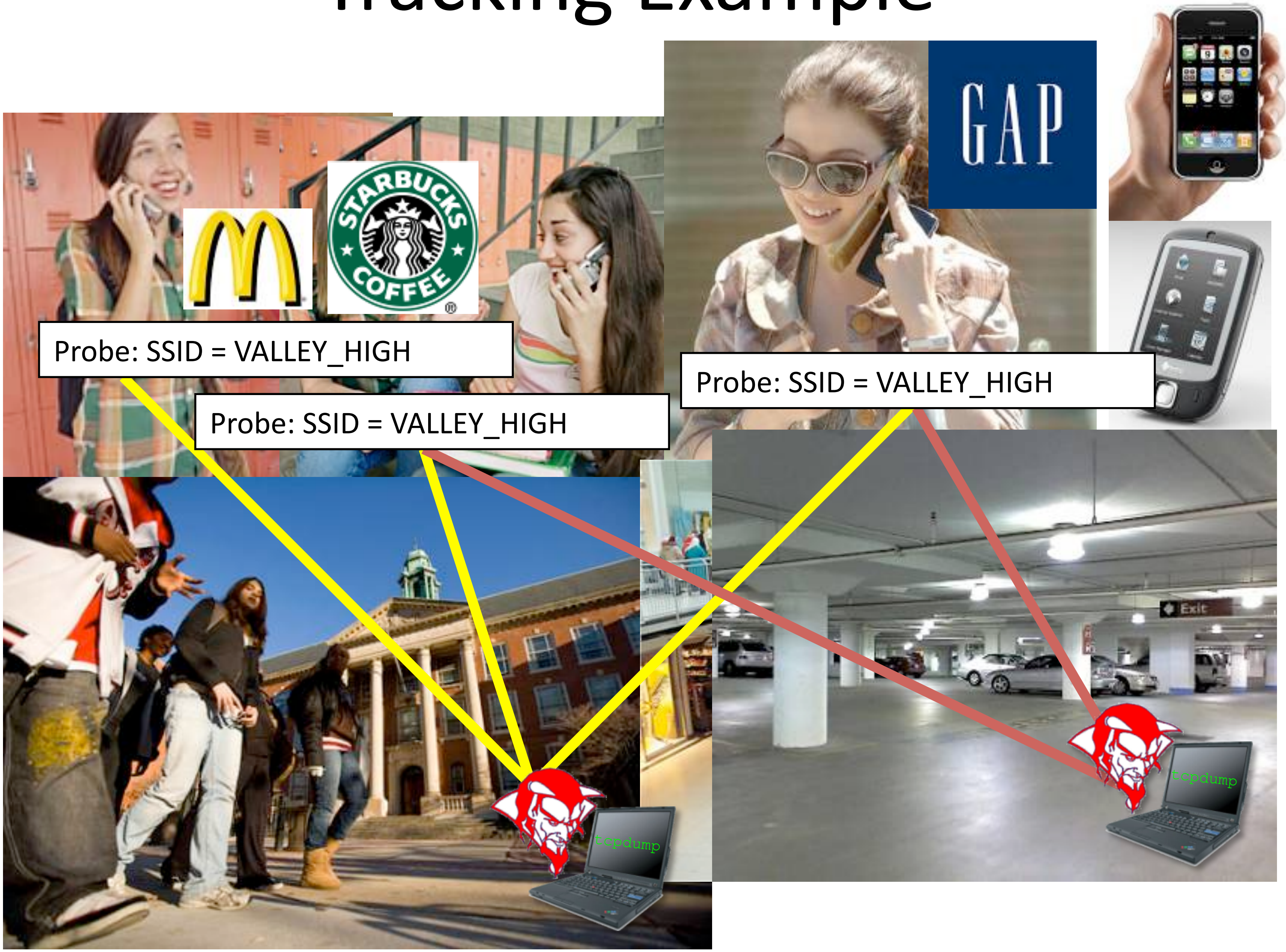
- EU directive 2006/24/EC: 3 year data retention
  - For ALL traffic, **requires EU ISPs to record**:
    - Sufficient information to identify endpoints (both legal entities and natural persons)
    - Session duration
    - … but not session contents
  - Make available to law enforcement
    - … but penalties for transfer or other access to data

- For info on US privacy on the net:
  - "privacy on the line" by W. Diffie and S. Landau

Threat 4: Untrustworthy Neighbors

- Can learn who you are and where you go, what you do, etc.

Tracking Example



Probe: SSID = VALLEY_HIGH

Probe: SSID = VALLEY_HIGH

Probe: SSID = VALLEY_HIGH

E.g., Bonjour

jill yetman’s iBook G4 [00:17:f2:c8:46:8e]._workstation._tcp.local

Gary Yngve’s MacBook Pro._postgresql._tcp.local

Benjamin Melton’s M

Chantri P

Marianne

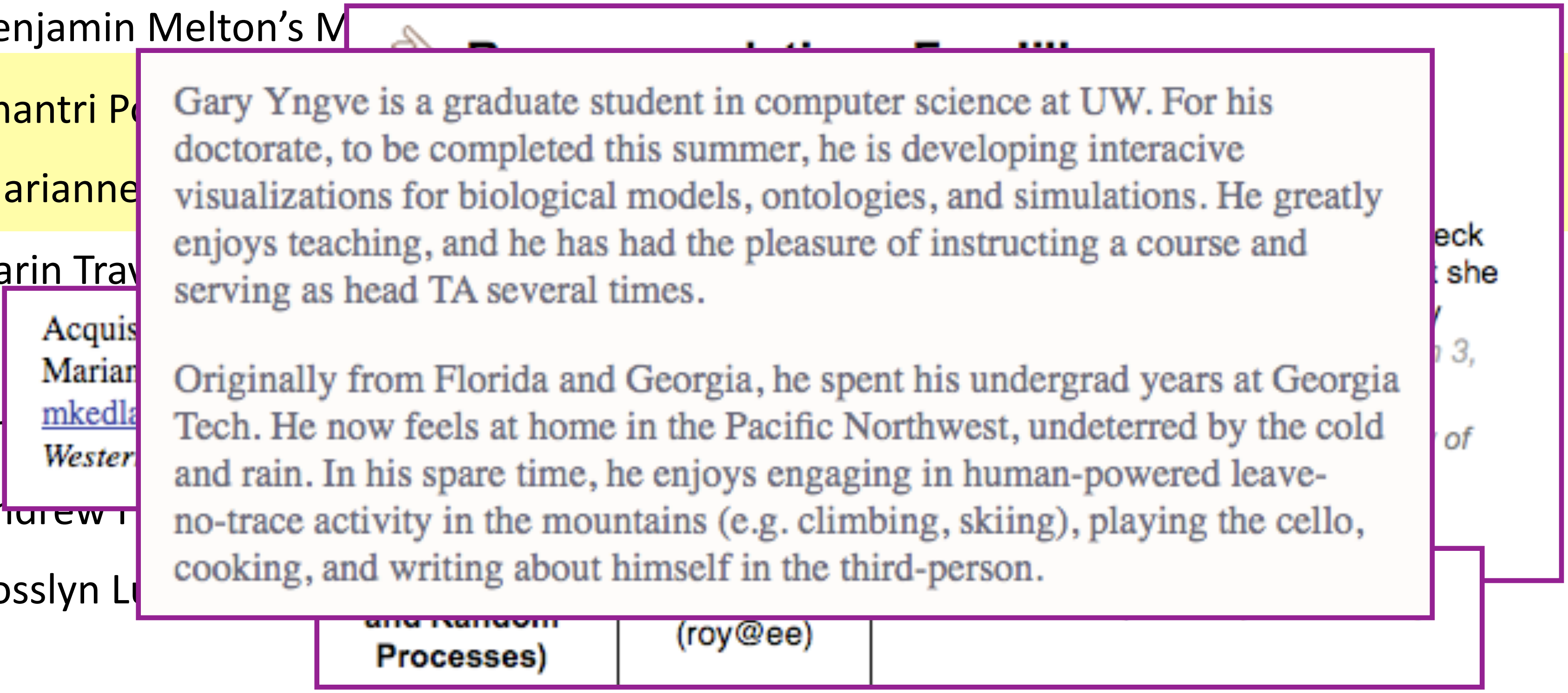
Darin Tra

Ja

Ch

Andrew

Rosslyn L



Gary Yngve is a graduate student in computer science at UW. For his doctorate, to be completed this summer, he is developing interacive visualizations for biological models, ontologies, and simulations. He greatly enjoys teaching, and he has had the pleasure of instructing a course and serving as head TA several times.

Originally from Florida and Georgia, he spent his undergrad years at Georgia Tech. He now feels at home in the Pacific Northwest, undeterred by the cold and rain. In his spare time, he enjoys engaging in human-powered leave-no-trace activity in the mountains (e.g. climbing, skiing), playing the cello, cooking, and writing about himself in the third-person.

Acquis
Marian
mkedla
Wester

eck
she
3,
of

Processes)

(roy@ee)

What are the defenses?

- Tor + Privoxy
- SlyFi

- Awareness?

TOR: For anonymous Web browsing

- Why?
  - Discuss health issues and financial matters anonymously, conceal interactions with gambling sites
  - Bypass Internet censorship in parts of the world
  - Law enforcement
- Two goals:
  - Hide user identity from target web site
  - Hide browsing pattern from employer or ISP

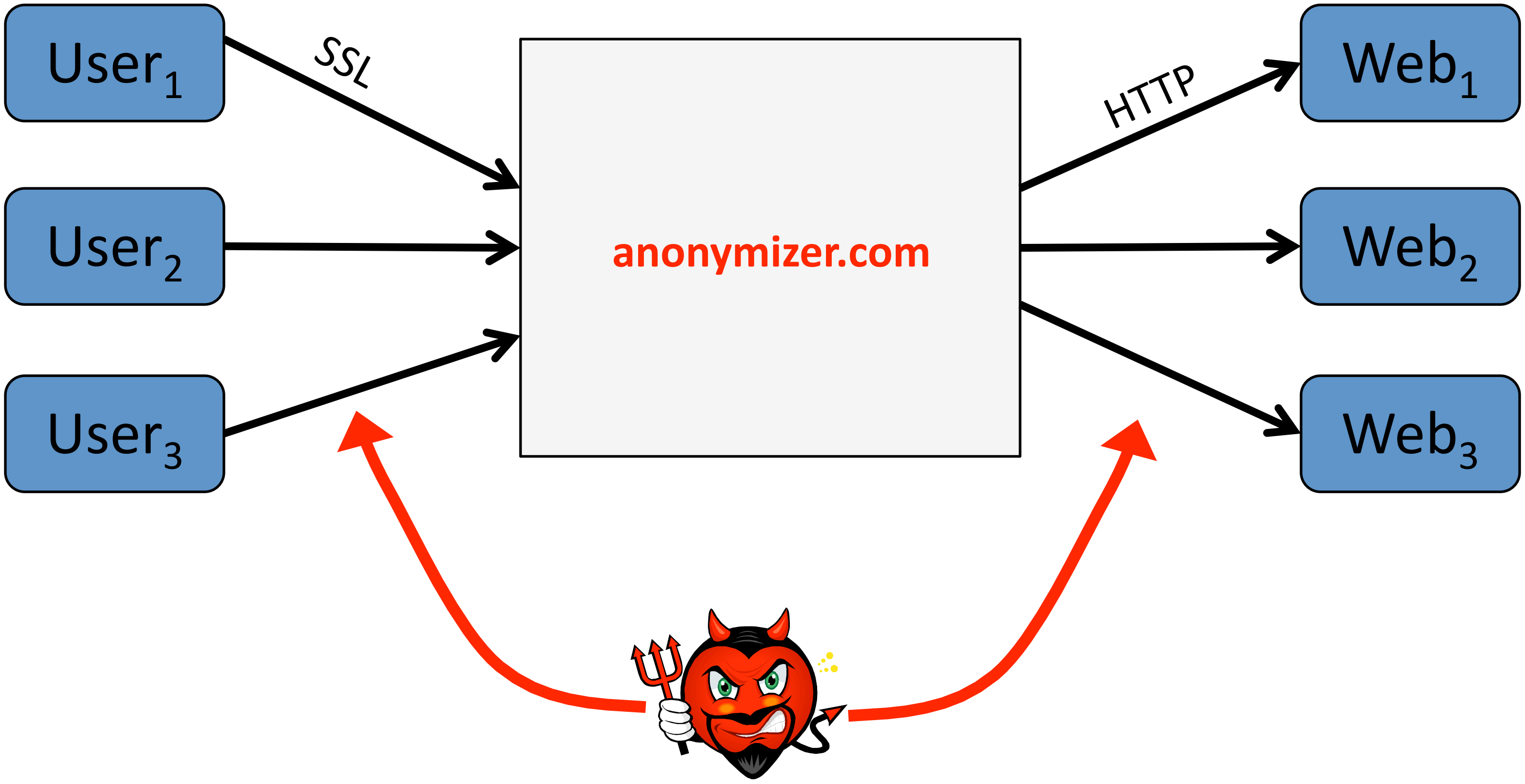
Part 1: network-layer privacy

Goals:

- Hide user's **IP address** from target web site
- Hide browsing destinations from network

1st attempt: anonymizing proxy

HTTPS:// anonymizer.com ? URL=target



Anonymizing proxy: security

- Monitoring ONE link: eavesdropper gets nothing
- Monitoring TWO links:
  - Eavesdropper can do traffic analysis
  - More difficult if lots of traffic through proxy

- Trust: proxy is a single point of failure
  - Can be corrupt or subpoenaed
    - Example: The Church of Scientology vs. anon.penet.fi

- Protocol issues:
  - Long-lived cookies make connections to site **linkable**

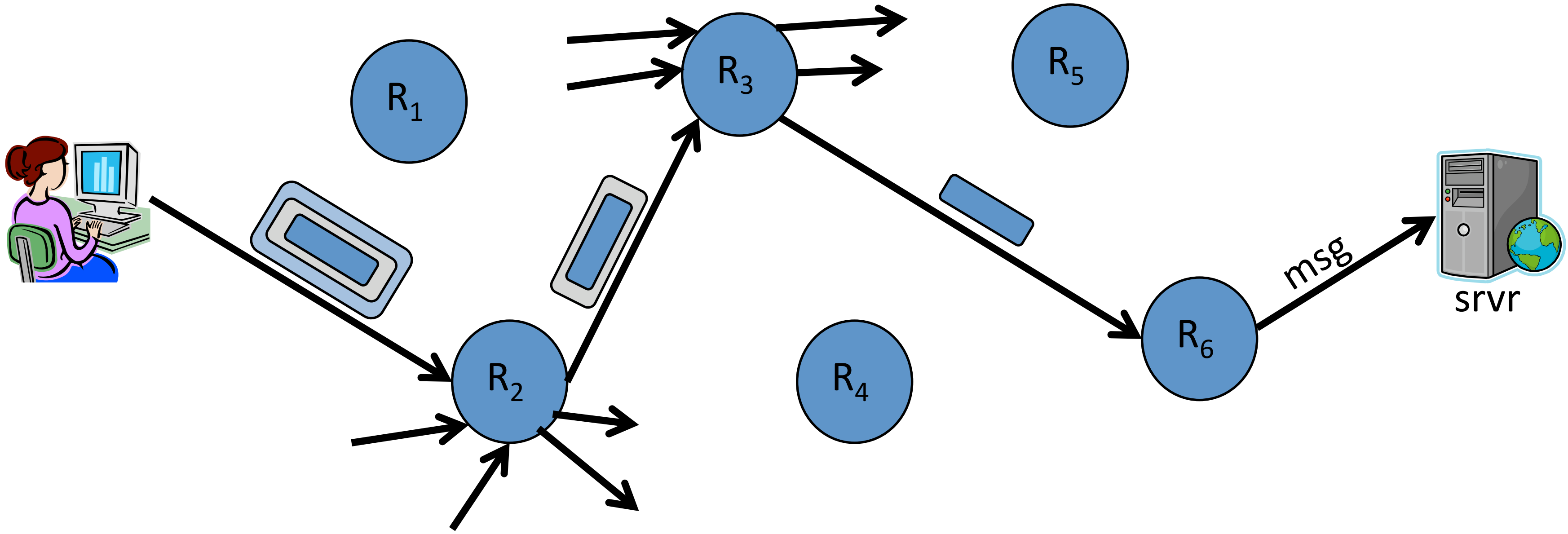
How proxy works

- Proxy rewrites all links in response from web site
  - Updated links point to anonymizer.com
    - Ensures all subsequent clicks are anonymized
- Proxy rewrites/removes cookies and some HTTP headers
- Proxy IP address:
  - if a single address, could be blocked by site or ISP
  - anonymizer.com consists of >20,000 addresses
    - Globally distributed, registered to multiple domains
    - Note: chinese firewall blocks ALL anonymizer.com addresses
- Other issues: attacks (click fraud) through proxy

2nd Attempt: MIX nets

Goal: no single point of failure

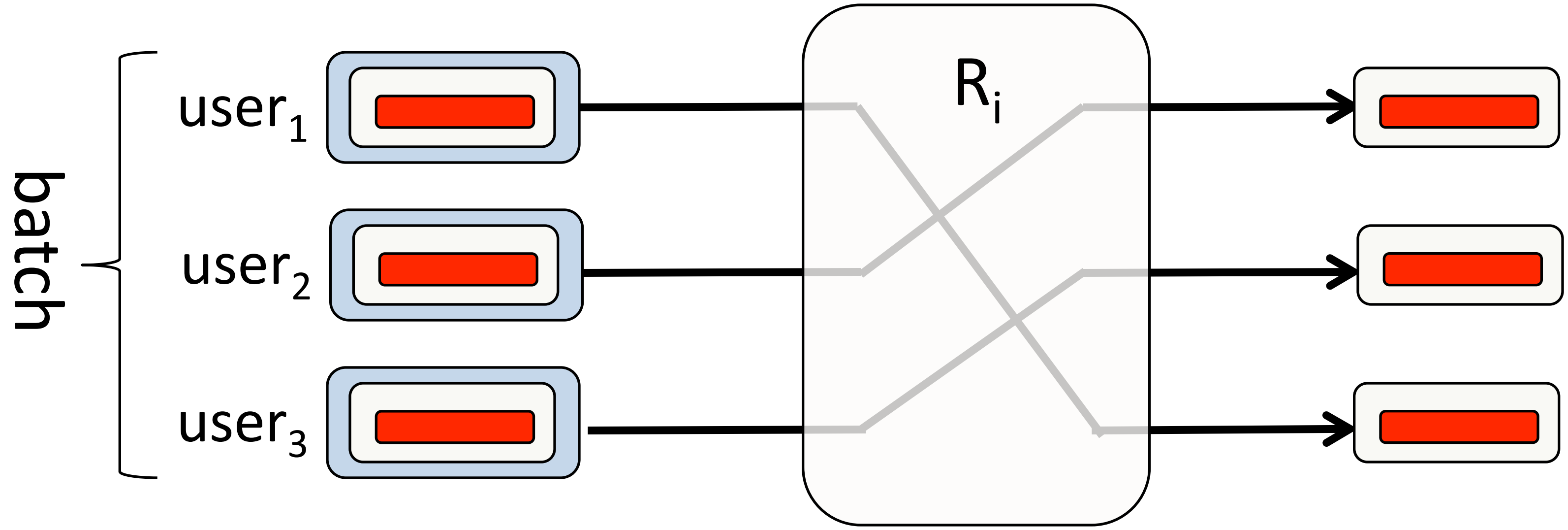
MIX nets [C’81]



- Every router has public/private key pair
  - Sender knows all public keys
- To send packet:
  - Pick random route: $R_2 \rightarrow R_3 \rightarrow R_6 \rightarrow$ srvr
  - Prepare **onion packet**:

packet = $E_{pk_2}(R_3, \; E_{pk_3}(R_6, \; E_{pk_6}(\text{srvr} , \text{msg})$

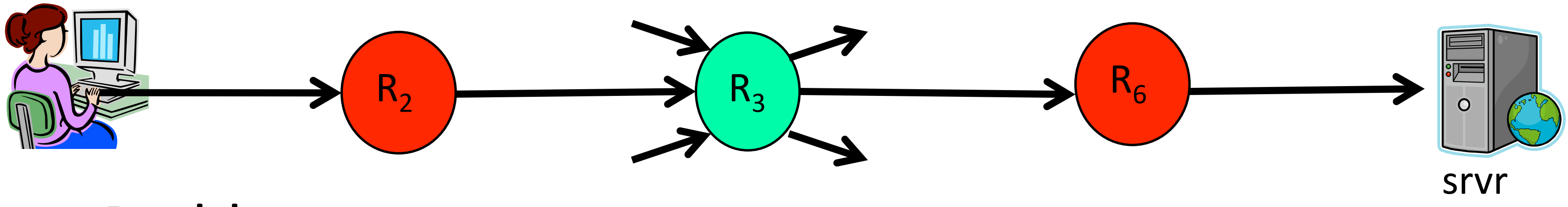
Eavesdropper’s view at a single MIX



- Eavesdropper observes incoming and outgoing traffic
- Crypto prevents linking input/output pairs
  - Assuming enough packets in incoming batch
  - If variable length packets then must pad all to max len
- Note: router is stateless

Performance

- Main benefit:
  - Privacy as long as **at least one** honest router on path



- <u>Problems</u>:
  - High latency (lots of public key ops)
    - Inappropriate for interactive sessions
    - May be OK for email (e.g. Babel system)
  - No forward security
- How does server respond?
  - hint: user includes “response onion” in forward packet

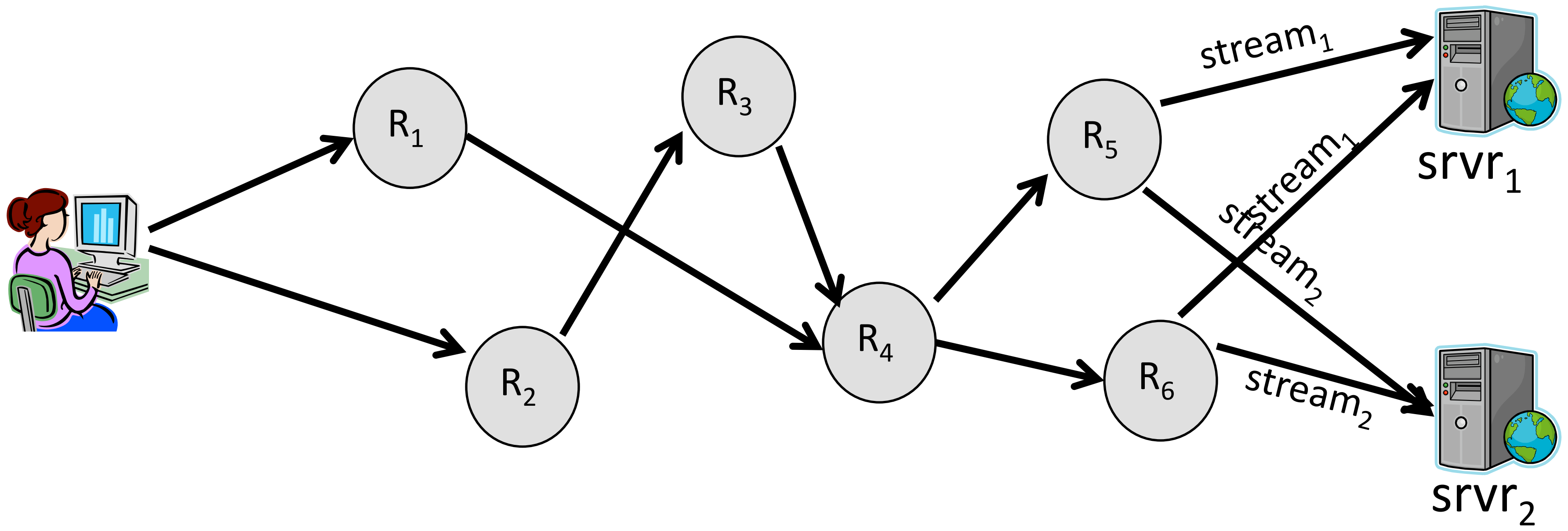
3rd Attempt: Tor MIX

circuit-based method

Goals: privacy as long as one honest router on path, and reasonable performance

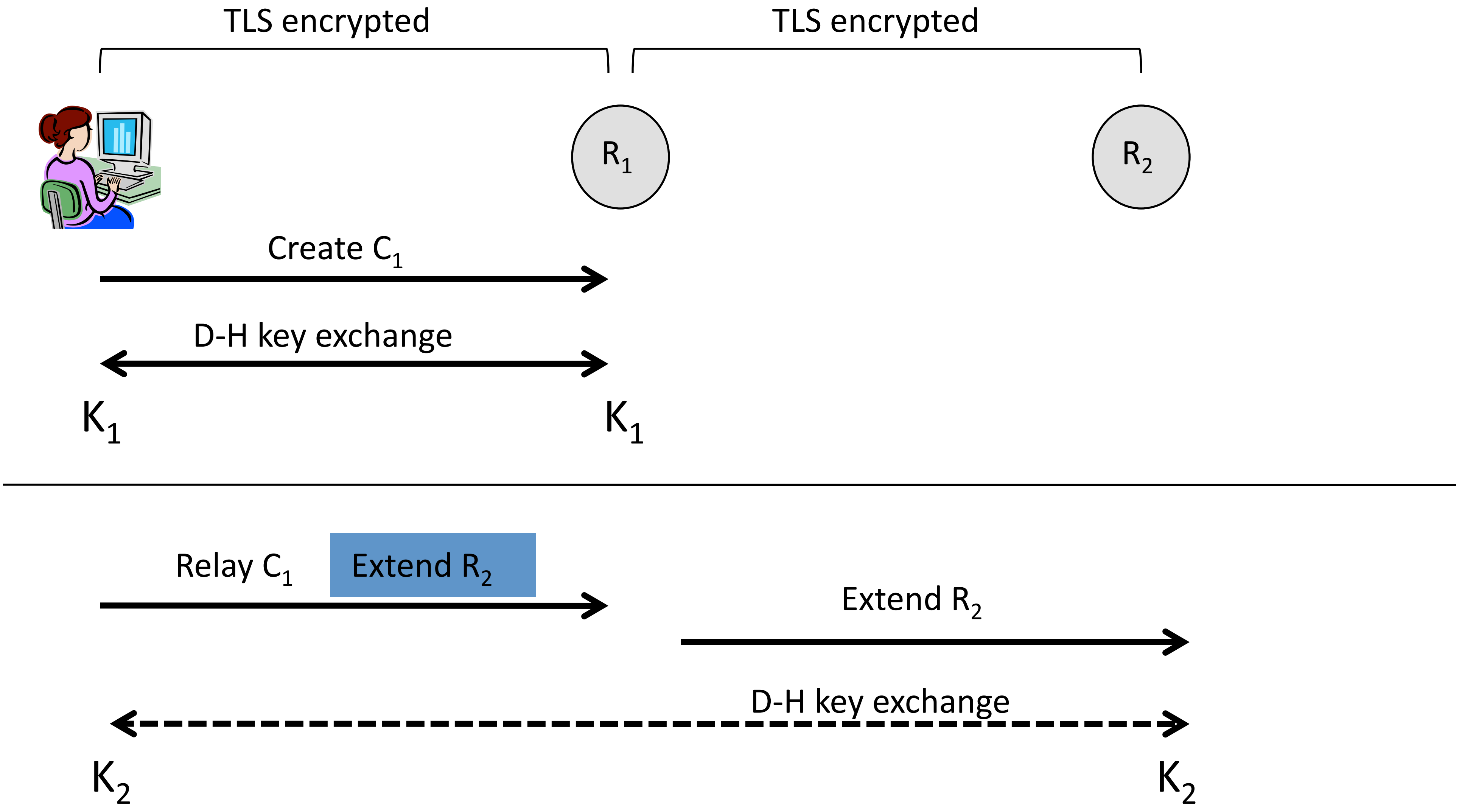
The Tor design

- Trusted directory contains list of Tor routers
- User's machine preemptively creates a circuit
  - Used for many TCP streams
  - New circuit is created once a minute



one minute later

Creating circuits



TLS encrypted

TLS encrypted

R_1

R_2

Create C_1

D-H key exchange

K_1

K_1

Relay C_1

Extend R_2

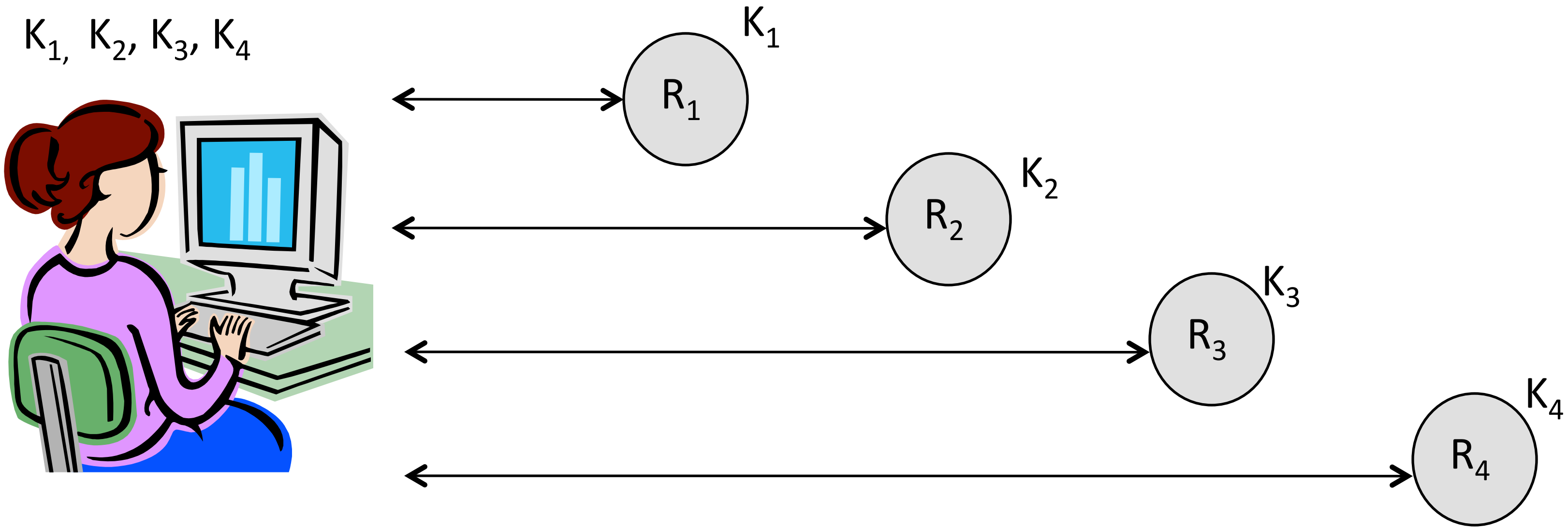
Extend R_2

D-H key exchange

K_2

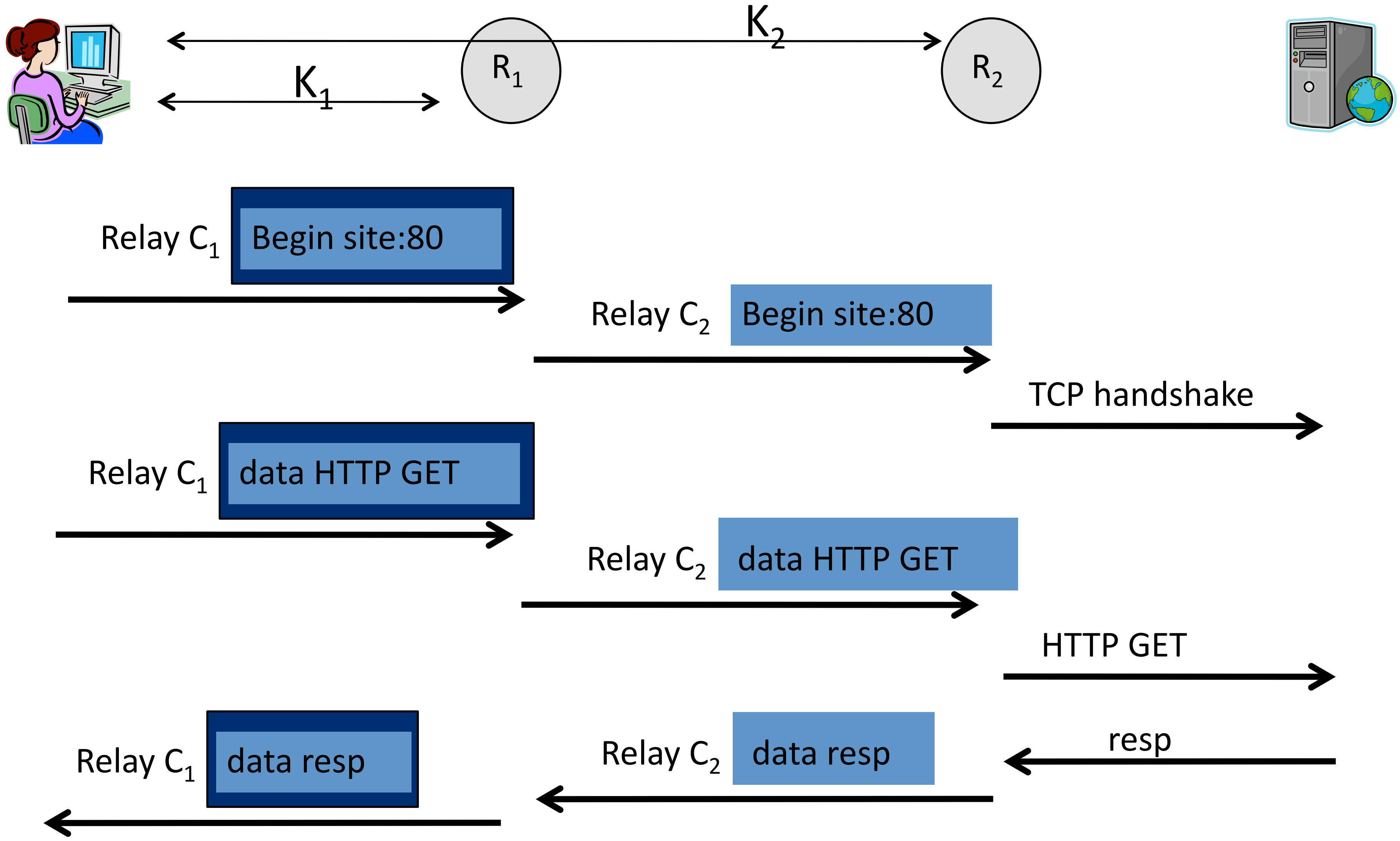
K_2

Once circuit is created



- User has shared key with each router in circuit
- Routers only know ID of successor and predecessor

Sending data



K_2

K_1

R_1

R_2

Relay C_1 Begin site:80

Relay C_2 Begin site:80

TCP handshake

Relay C_1 data HTTP GET

Relay C_2 data HTTP GET

HTTP GET

Relay C_1 data resp

Relay C_2 data resp

resp

Properties

- Performance:
  - Fast connection time: circuit is pre-established
  - Traffic encrypted with AES: no pub-key on traffic
- Tor crypto:
  - provides end-to-end integrity for traffic
  - Forward secrecy via TLS
- Downside:
  - Routers must maintain state per circuit
  - Each router can link multiple streams via `CircuitID`
    - all steams in one minute interval share same `CircuitID`

Privoxy

- Tor only provides network level privacy
  - No application-level privacy
    - e.g. mail progs add "From: email-addr" to outgoing mail

- Privoxy:
  - Web proxy for browser-level privacy
  - Removes/modifies cookies
  - Other web page filtering

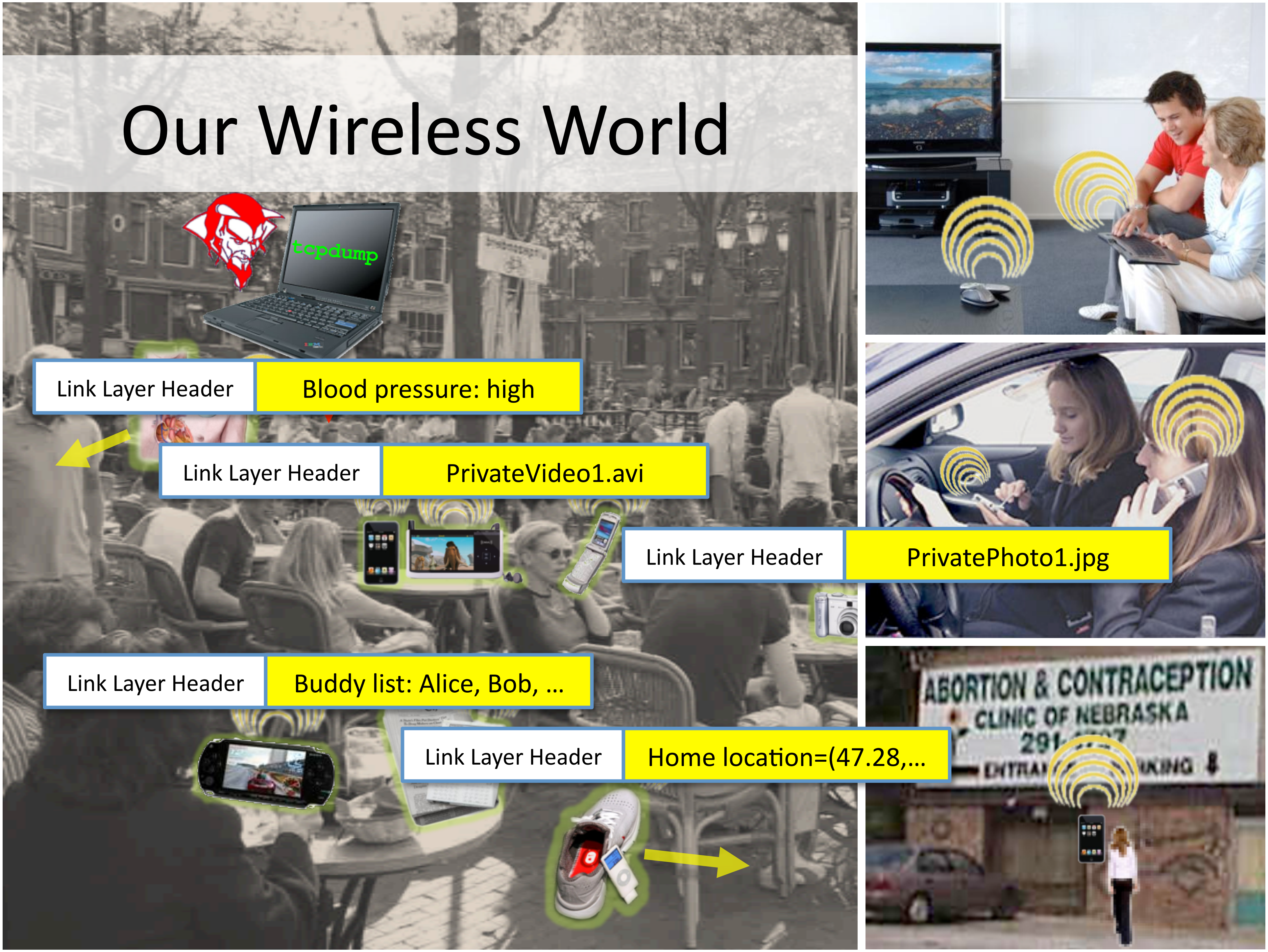
SlyFi: A link layer with better privacy

- Why?
  - Easy for eavesdropper to track and profile users via their wireless transmissions
- Two goals:
  - Conceal all identifiers from 3rd parties
  - Make the protocol efficient

  - Who: Ben Greenstein, Damon McCoy, Jeffrey Pang, Tadayoshi Kohno, Srinivasan Seshan, and David Wetherall

  - The paper itself: http://www.seattle.intel-research.net/pubs/mobisys08-slyfi.pdf

Our Wireless World



Link Layer Header | Blood pressure: high

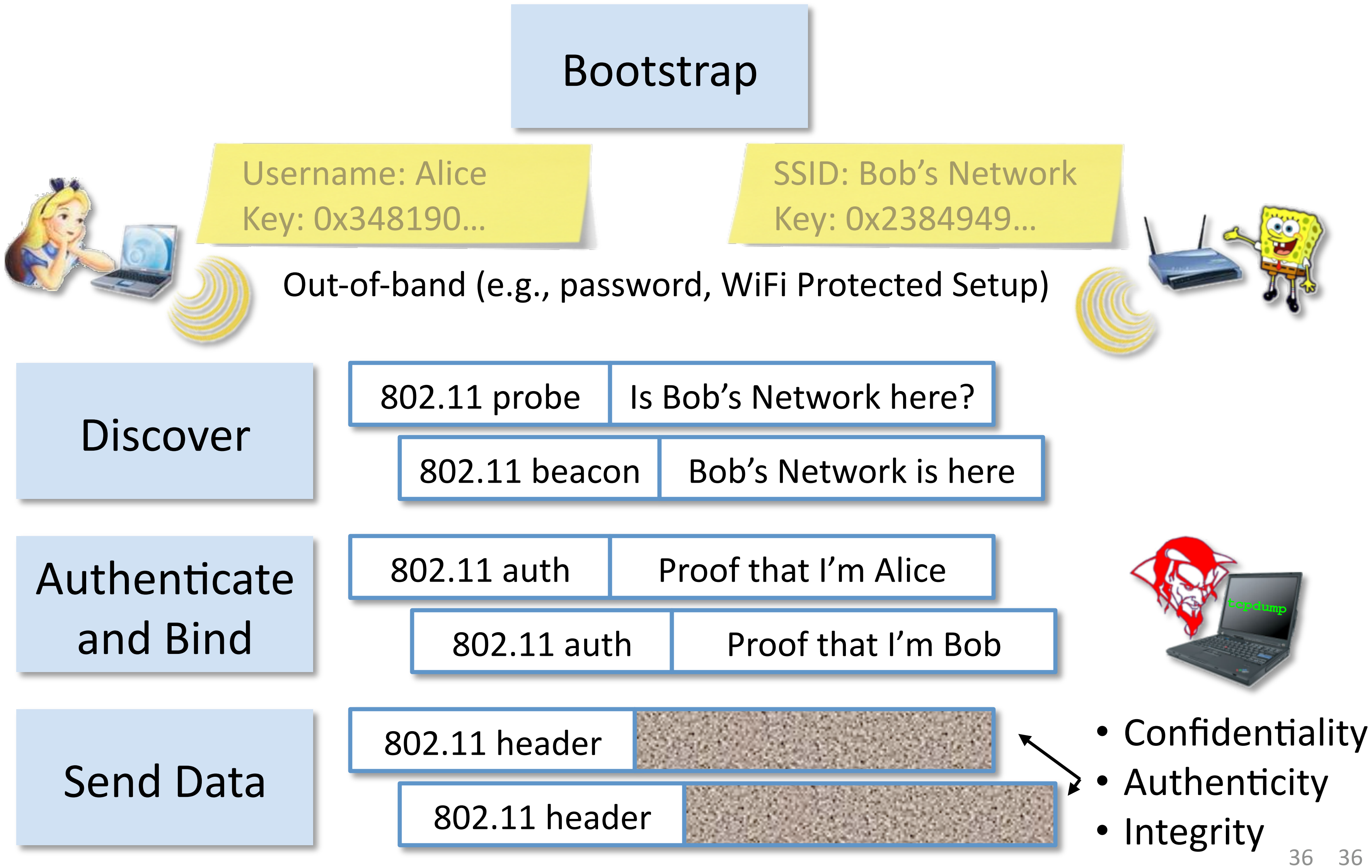
Link Layer Header | PrivateVideo1.avi

Link Layer Header | PrivatePhoto1.jpg

Link Layer Header | Buddy list: Alice, Bob, …

Link Layer Header | Home location=(47.28,…

Best Security Practices



Bootstrap

Username: Alice
Key: 0x348190…

SSID: Bob's Network
Key: 0x2384949…

Out-of-band (e.g., password, WiFi Protected Setup)

Discover

802.11 probe	Is Bob's Network here?
802.11 beacon	Bob's Network is here

Authenticate and Bind

802.11 auth	Proof that I'm Alice
802.11 auth	Proof that I'm Bob

Send Data

802.11 header	
802.11 header	

- Confidentiality
- Authenticity
- Integrity

Privacy Problems Remain

Many exposed bits are (or can be used as) identifiers that are linked over time

Secret: 0x2384949…

Secret: 0x348190…

Discover

- 802.11 probe | Is Bob's Network here?
- 802.11 beacon | Bob's Network is here

Authenticate and Bind

- 802.11 auth | Proof that I'm Alice
- 802.11 auth | Proof that I'm Bob

Send Data

- MAC addr, seqno, …
- MAC addr, seqno, …

- Confidentiality
- Authenticity
- Integrity

Problem: Long-Term Linking

802.11 beacon

Alice’s iPod is here

MAC: 12:34:56:78:90:ab

802.11 beacon

Alice’s iPod is here

MAC: 12:34:56:78:90:ab

802.11 probe

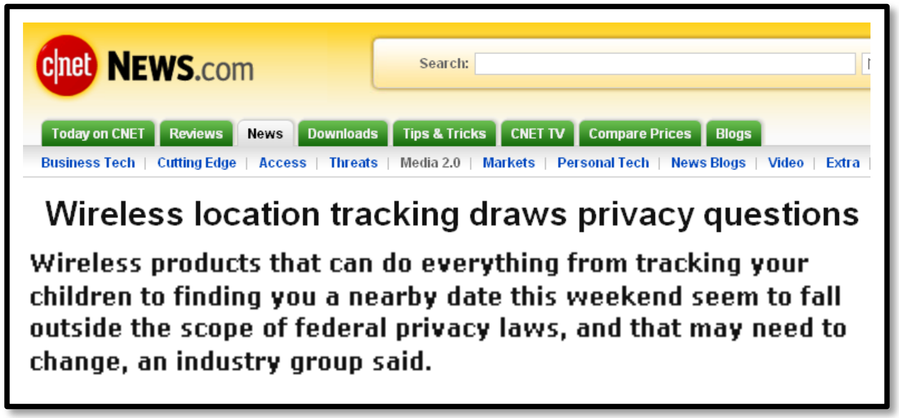
Is Alice’s iPod here?

Easy to identify and relate devices over time

Problem: Long-Term Linking

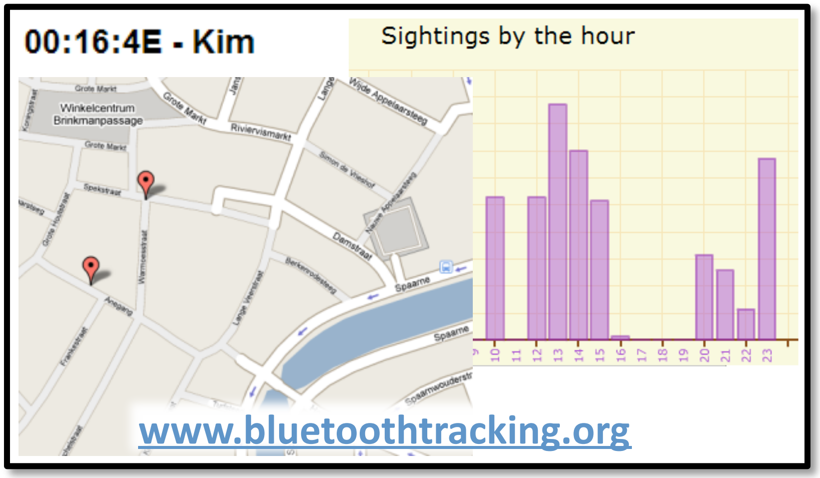
Linking enables location tracking, user profiling, inventorying, relationship profiling, …

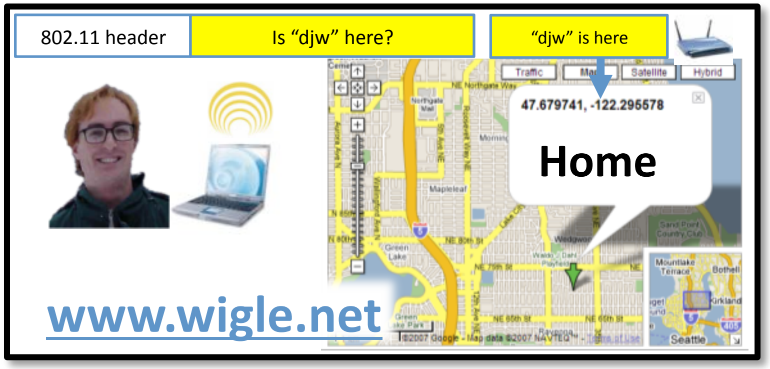
[Greenstein, *HotOS* ’07; Jiang, *MobiSys* ’07; Pang, *MobiCom* ’07, *HotNets* ’07]

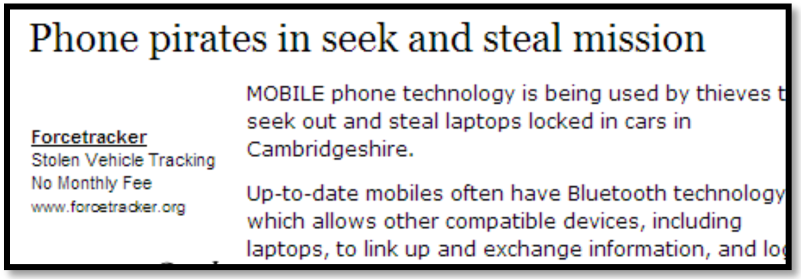


Wireless location tracking draws privacy questions

Wireless products that can do everything from tracking your children to finding you a nearby date this weekend seem to fall outside the scope of federal privacy laws, and that may need to change, an industry group said.





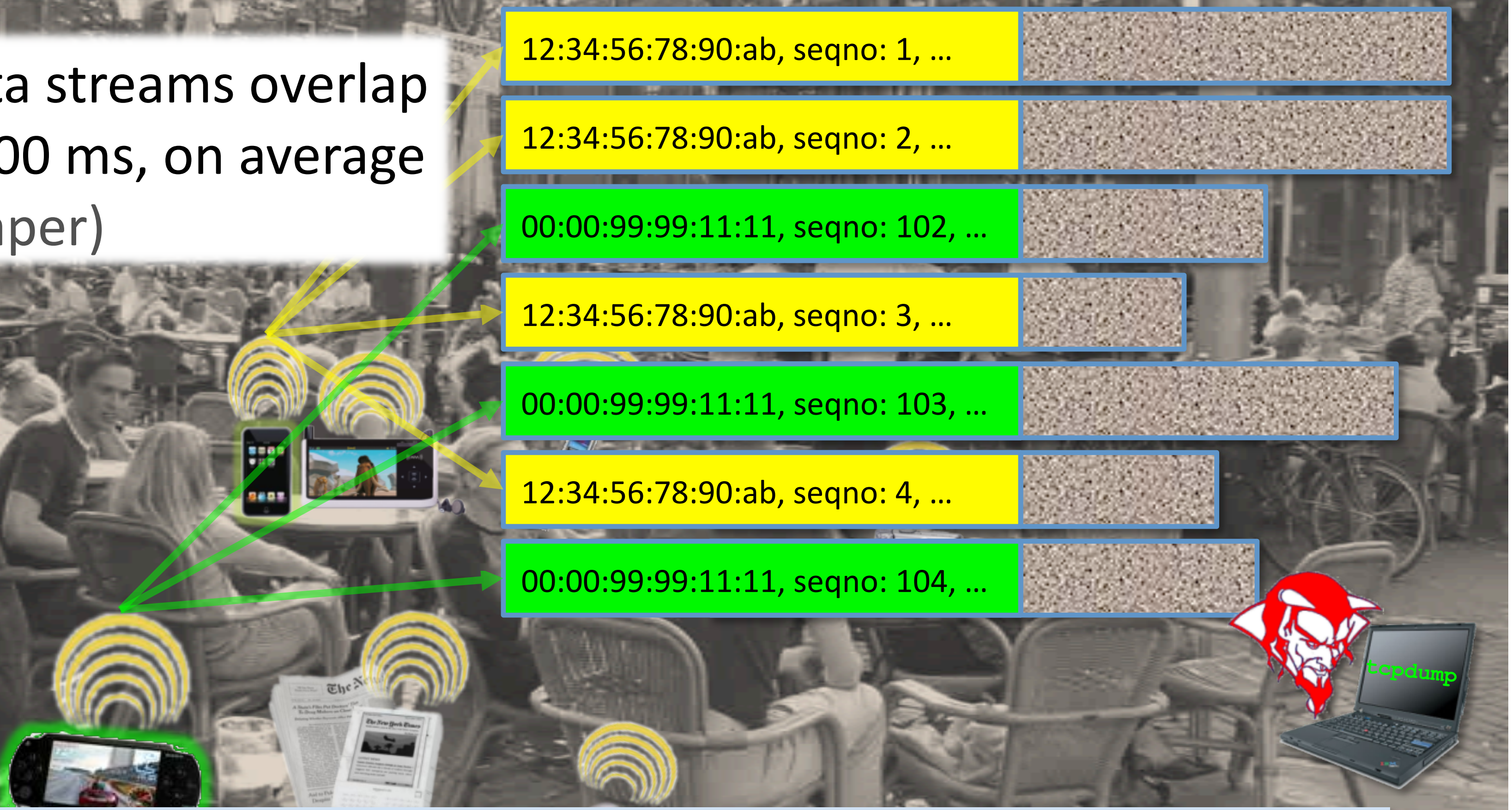

Phone pirates in seek and steal mission

MOBILE phone technology is being used by thieves t seek out and steal laptops locked in cars in Cambridgeshire.

Up-to-date mobiles often have Bluetooth technology which allows other compatible devices, including laptops, to link up and exchange information, and lo

Problem: Short-Term Linking

3-9 data streams overlap each 100 ms, on average (see paper)



12:34:56:78:90:ab, seqno: 1, …

12:34:56:78:90:ab, seqno: 2, …

00:00:99:99:11:11, seqno: 102, …

12:34:56:78:90:ab, seqno: 3, …

00:00:99:99:11:11, seqno: 103, …

12:34:56:78:90:ab, seqno: 4, …

00:00:99:99:11:11, seqno: 104, …

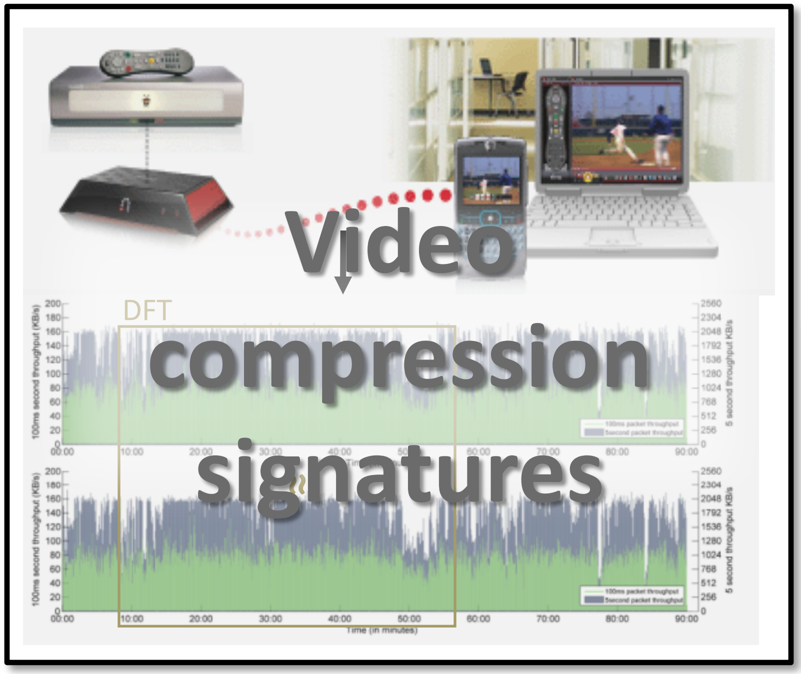
Easy to isolate distinct packet streams

Problem: Short-Term Linking

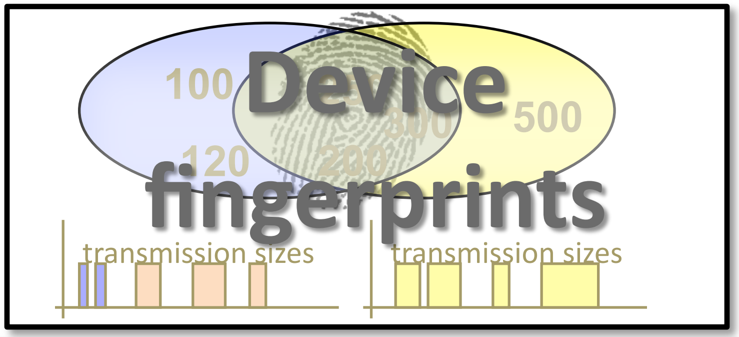
Isolated data streams are more susceptible to side-channel analysis on packet sizes and timing

- Exposes keystrokes, VoIP calls, webpages, movies, …

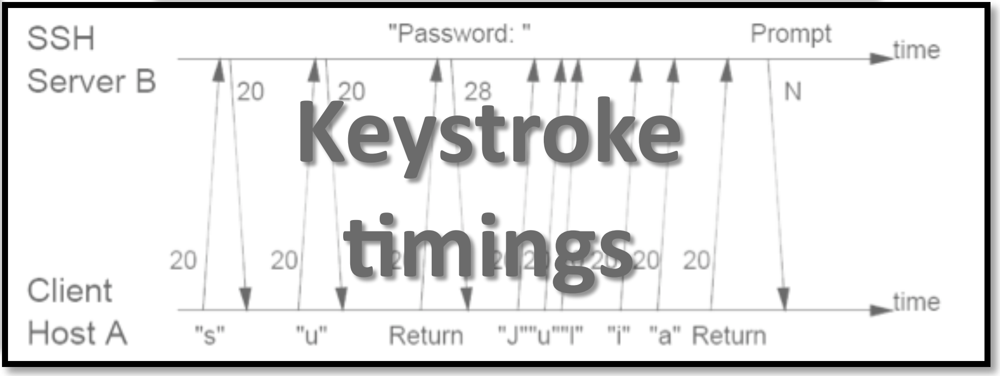
[Liberatore, *CCS* ‘06; Pang, *MobiCom* ’07; Saponas, *Usenix Security* ’07; Song, *Usenix Security* ‘01; Wright, *IEEE S&P ‘08*; Wright, *Usenix Security* ‘07]



Video compression signatures



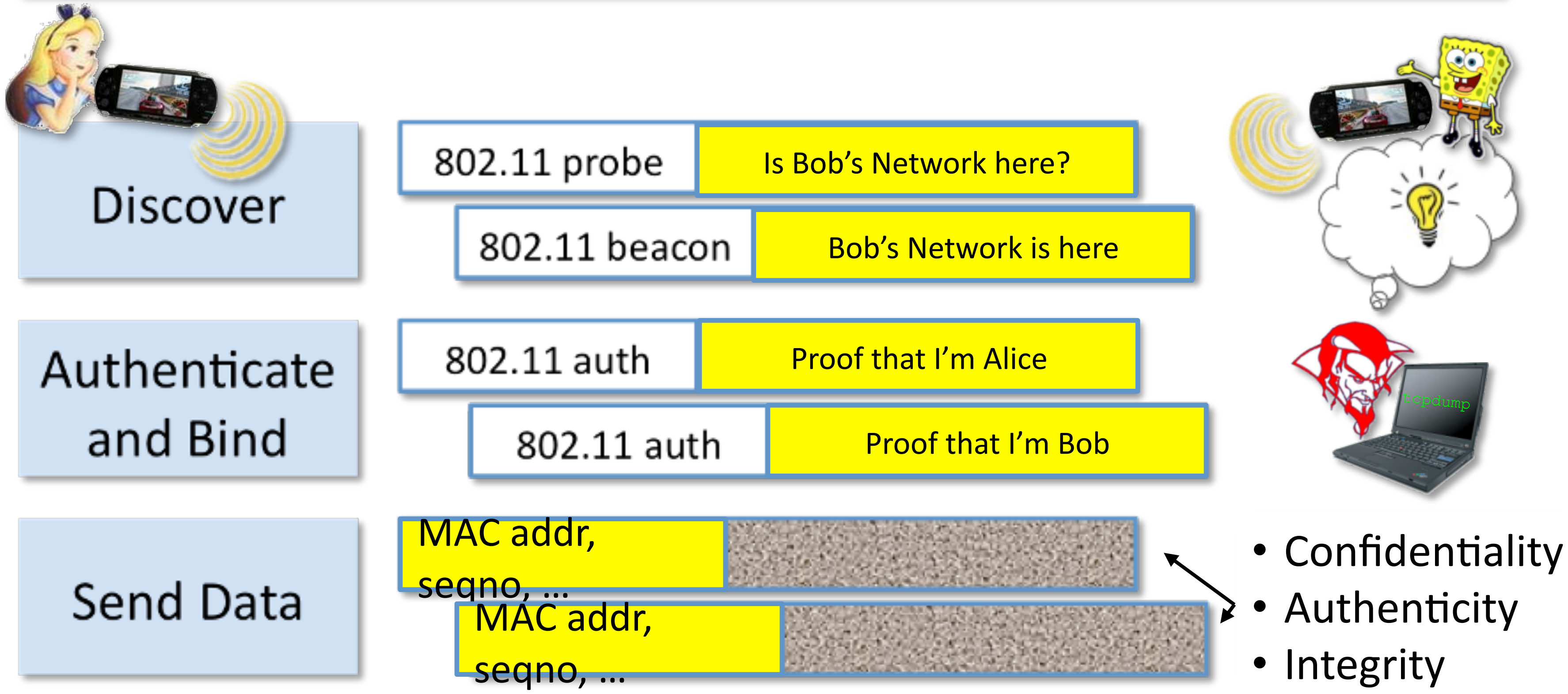
Device fingerprints



Keystroke timings

Key technical problem to solve

Many exposed bits are (or can be used as) identifiers that are linked over time



Discover

802.11 probe | Is Bob's Network here?

802.11 beacon | Bob's Network is here

Authenticate and Bind

802.11 auth | Proof that I'm Alice

802.11 auth | Proof that I'm Bob

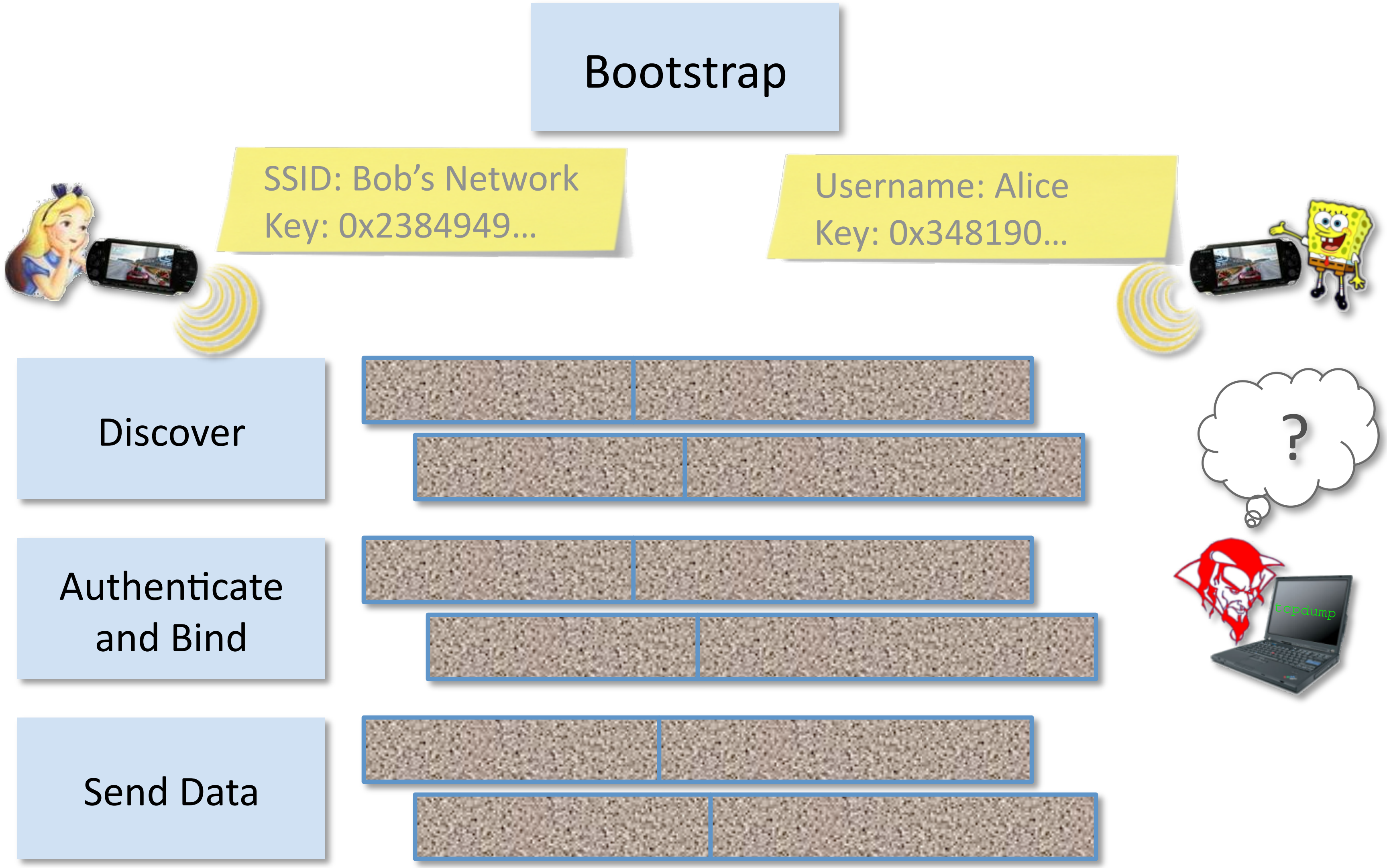
Send Data

MAC addr, seqno, …

MAC addr, seqno, …

- Confidentiality
- Authenticity
- Integrity

Goal: all bits appear random to outsiders



Bootstrap

SSID: Bob's Network
Key: 0x2384949…

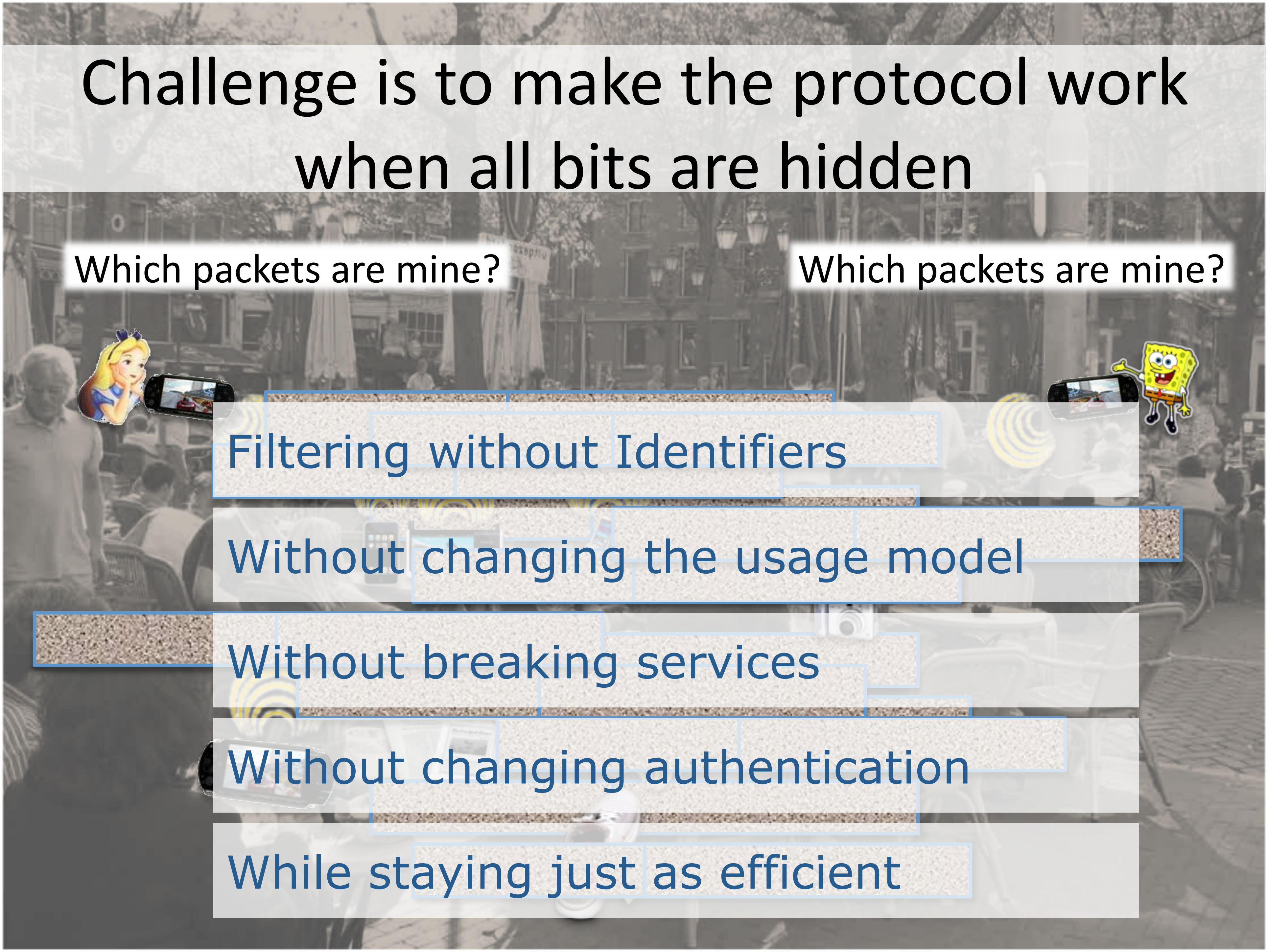
Username: Alice
Key: 0x348190…

Discover

Authenticate and Bind

Send Data

Challenge is to make the protocol work when all bits are hidden



Filtering without Identifiers

Without changing the usage model

Without breaking services

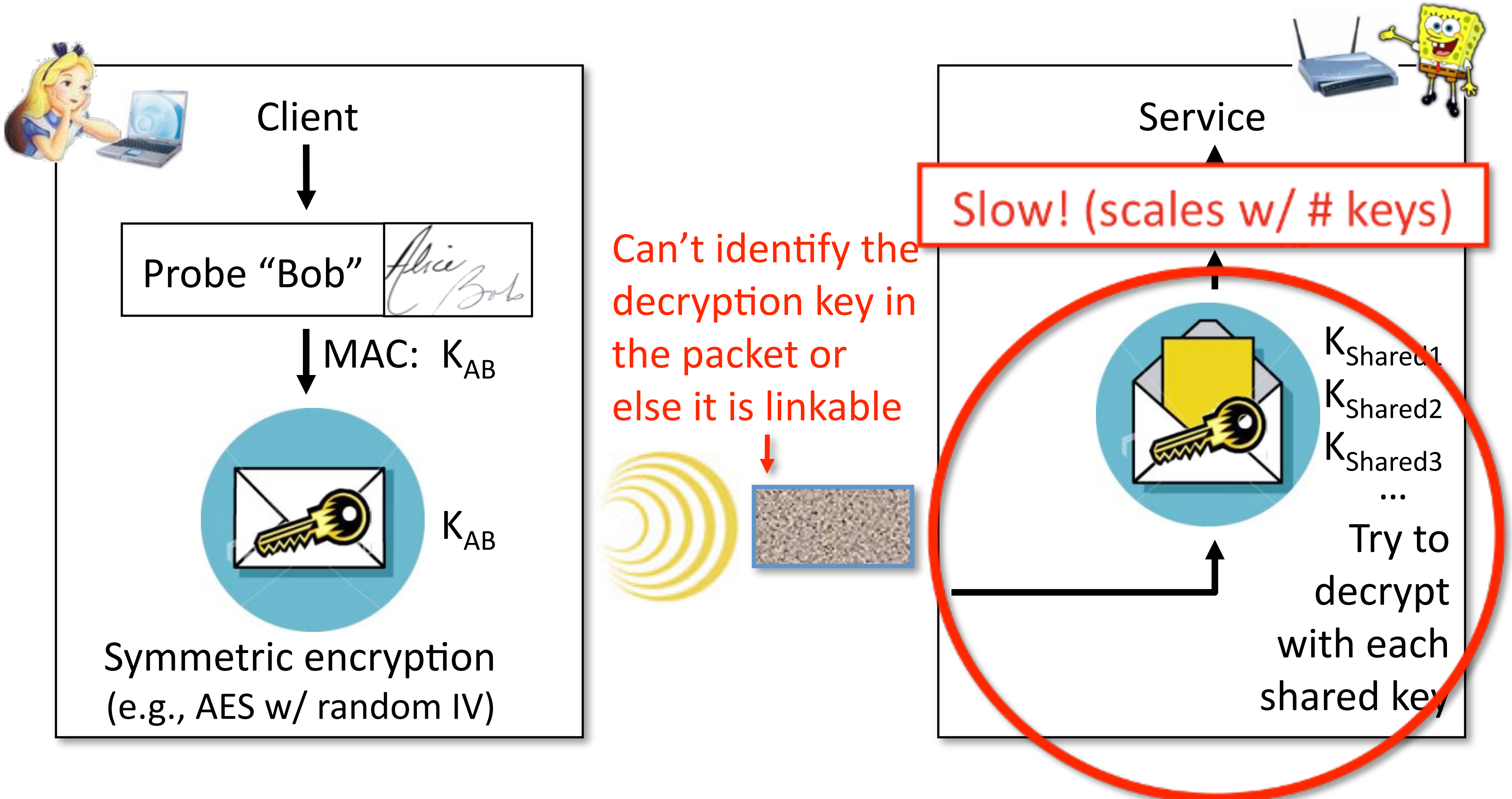
Without changing authentication

While staying just as efficient

Straw man: MAC Pseudonyms

- **Idea**: change MAC address periodically
  - Per session or when idle [Gruteser ’05, Jiang ‘07]
- Other fields remain (e.g., in discovery/binding)
  - No mechanism for data authentication/encryption
  - Doesn’t hide network names during discovery or credentials during authentication
- Pseudonyms are linkable in the short-term
  - Same MAC must be used for each association
  - Data streams still vulnerable to side-channel leaks

Naïve approach (symmetric encryption of all bits) is slow



Client

Probe "Bob"

Alice Bob

MAC: K_{AB}

K_{AB}

Symmetric encryption
(e.g., AES w/ random IV)

Can't identify the decryption key in the packet or else it is linkable

Service

Slow! (scales w/ # keys)

$K_{Shared1}$
$K_{Shared2}$
$K_{Shared3}$
…

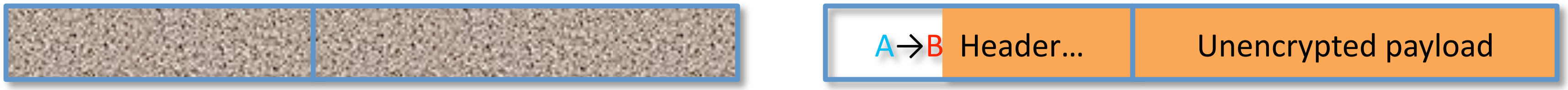
Try to decrypt with each shared key

Different symmetric key per potential sender

Design Requirement: add privacy to security without breaking anything else

- When *A* generates *Message* to *B*, she sends:

PrivateMsg = F(*A*, *B*, *Message*)

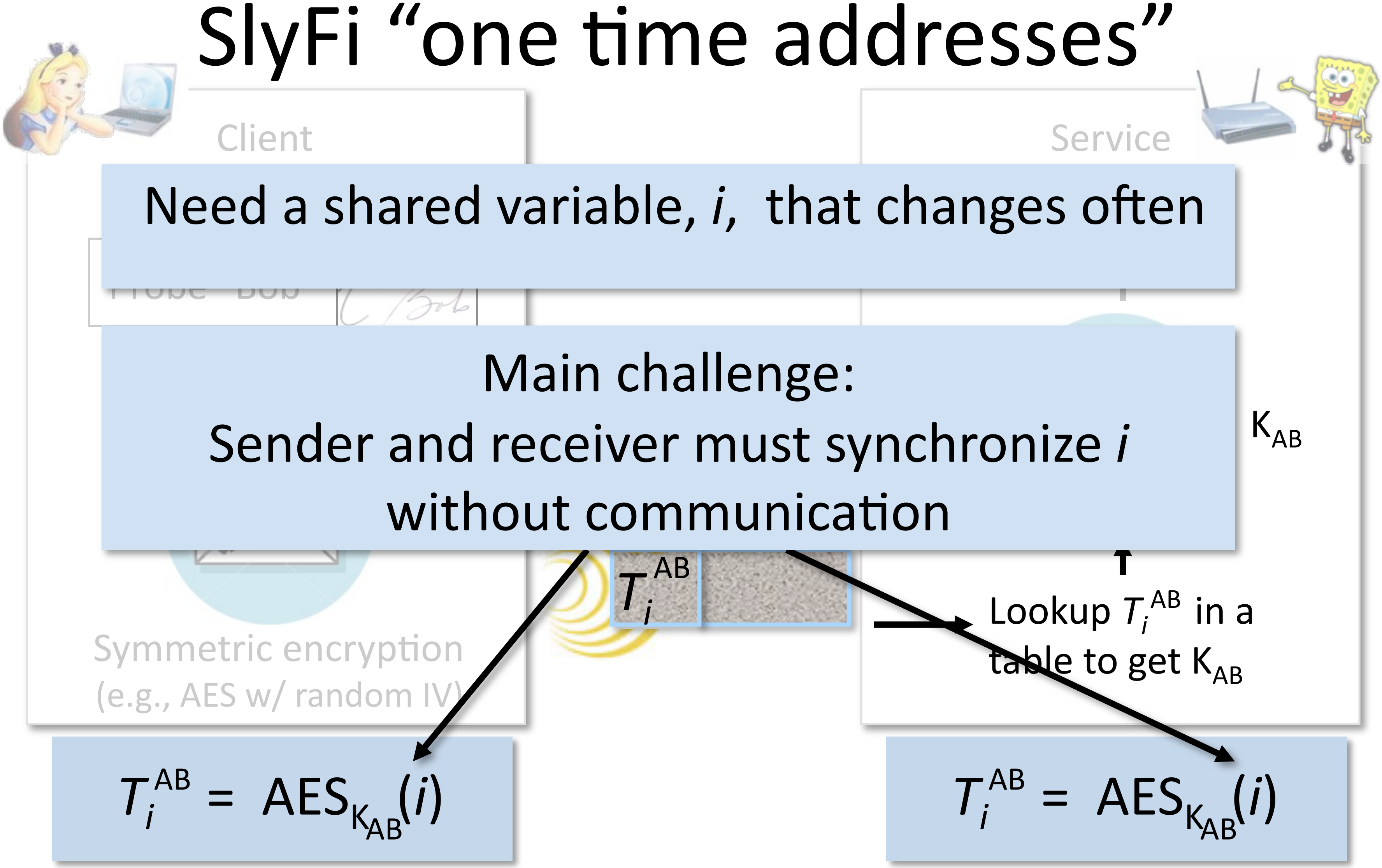


- Where F has these properties:
  - **Confidentiality**: Only *A* and *B* can determine *Message.*
  - **Authenticity**: *B* can verify *A* created *PrivateMsg.*
  - **Integrity**: *B* can verify *Message* not modified
  - **Unlinkability**: Only *A* and *B* can link *PrivateMsgs* to same sender or receiver
  - **Efficiency**: *B* can process *PrivateMsgs* as fast as he can receive them

SlyFi insight: split encryption to provide “one-time addresses”

- Symmetric key almost works, but tension between:
  - Unlinkability: can’t expose the identity of the key
  - Efficiency: need to identify the key to avoid trying all keys
- **Idea**: Split the encryption to identify the key in an unlinkable way. Provides “one-time addresses” that can be pre-computed for efficient matching
- Approach:
  - Sender **A** and receiver **B** agree on tokens: T_1^{AB}, T_2^{AB}, T_3^{AB}, …
  - **A** attaches T_i^{AB} to encrypted packet for **B**

SlyFi “one time addresses”



Need a shared variable, *i*, that changes often

Main challenge:

Sender and receiver must synchronize *i* without communication

$T_i^{AB} = \text{AES}_{K_{AB}}(i)$

$T_i^{AB} = \text{AES}_{K_{AB}}(i)$

Data Transport

= transmission #

- Only sent over established connections
- Expect messages to be delivered

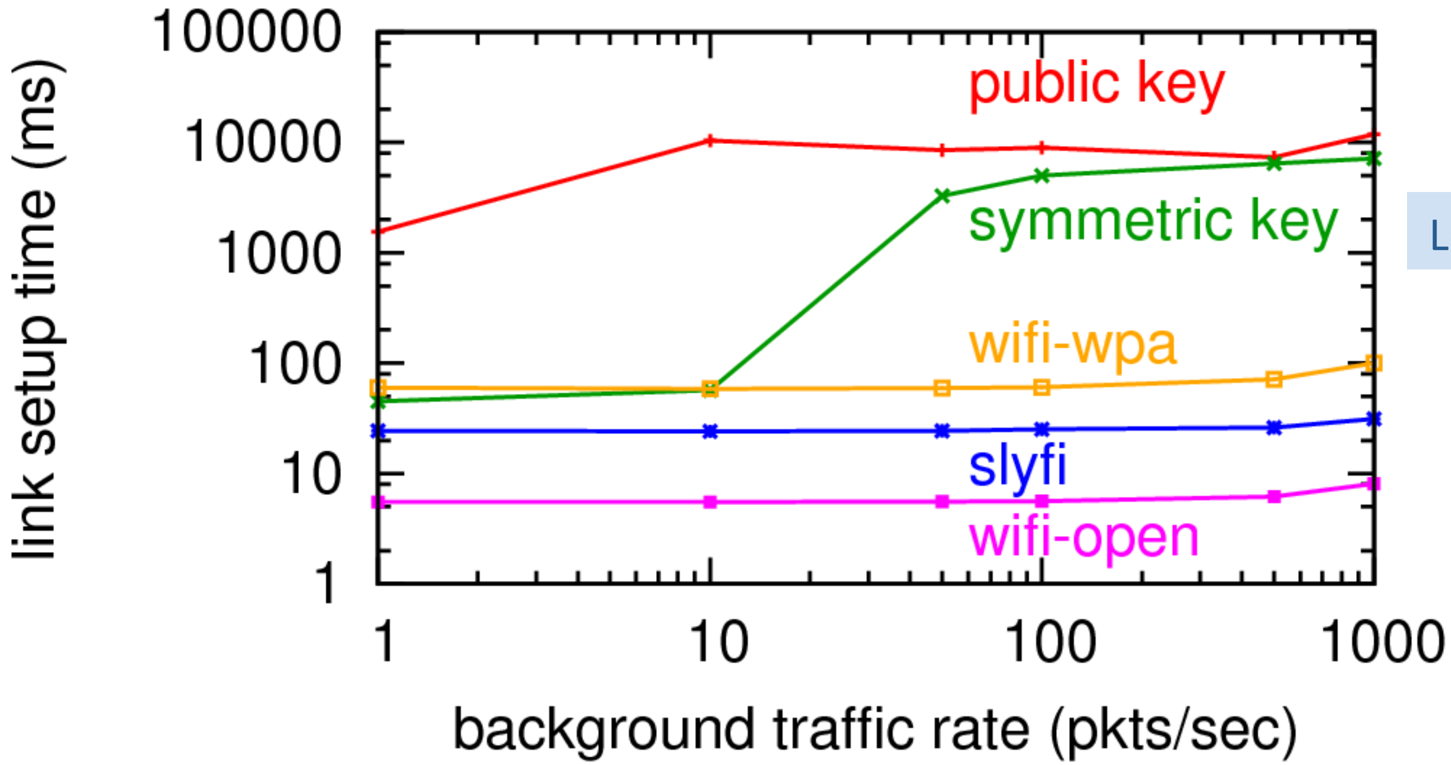
Discovery and Binding

$= \lfloor \text{current time}/5 \text{ min} \rfloor$

- **Infrequent**: sent when trying to associate
- **Narrow interface**: single application, few side-channels
- Linkability OK at short timescales

- On receipt of $T_i^{AB} = \mathrm{AES}_{K_{AB}}(i)$, receiver computes T_{i+1}^{AB}
- Handling message loss or clock skew:
  - On receipt of T_i^{AB} save $T_{i+1}^{AB}, \ldots, T_{i+k}^{AB}$ in table
  - Tolerates *k* consecutive losses or skew of 5 * *k* minutes
  - No loss $\Rightarrow$ compute one token per reception

Prototype Discovery/Binding Time



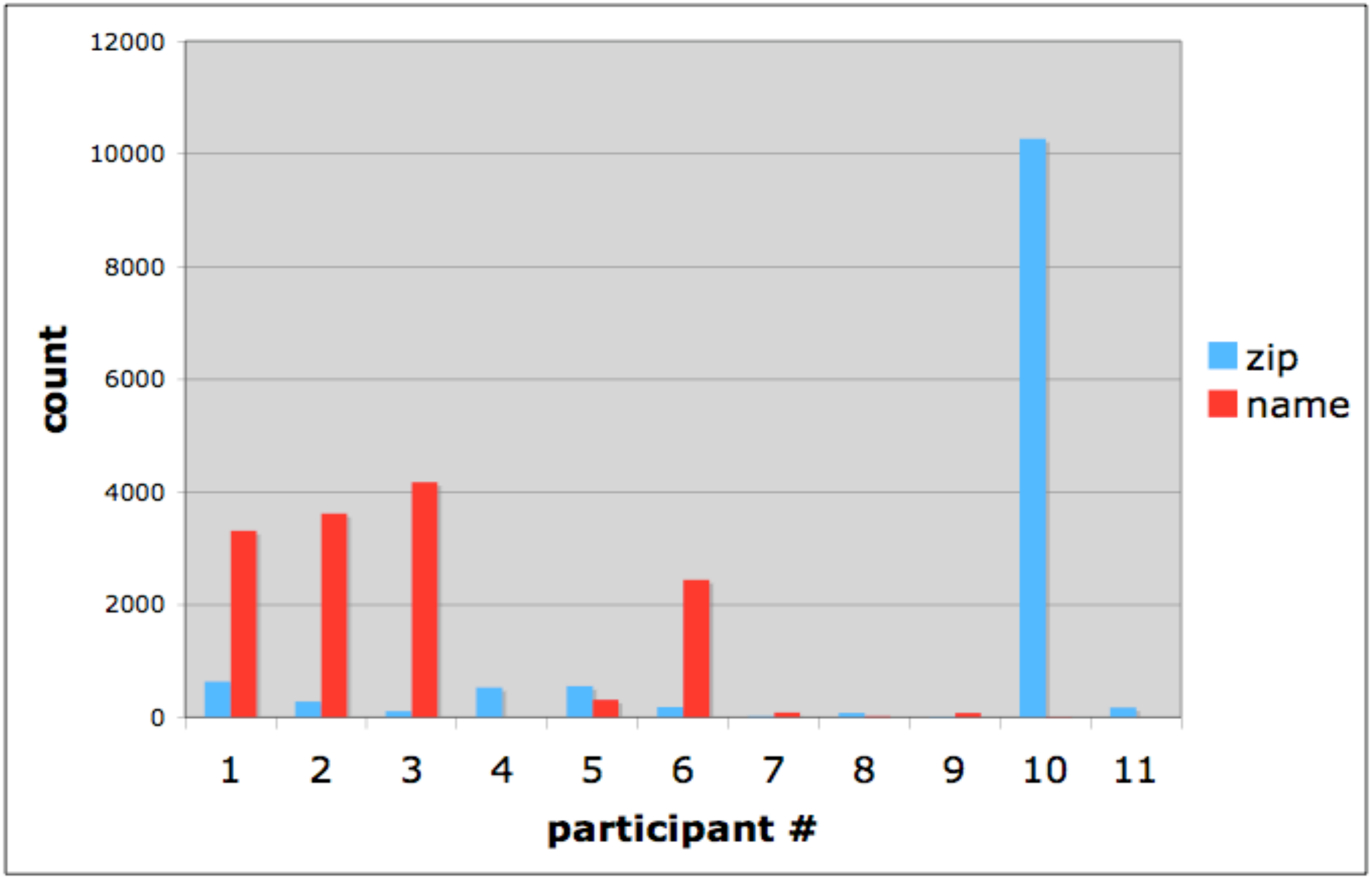
Lower = Better

SlyFi link setup has less overhead than WPA

Solution Summary

	Confidentiality	Authenticity	Integrity	Unlinkability	Efficiency
802.11 WPA (today)	Only Data Payload	Only Data Payload	Only Data Payload	🚫	✓
MAC Pseudonyms	🚫	🚫	🚫	Long Term	✓
Naïve Symmetric Key	✓	✓	✓	✓	🚫
SlyFi: Discovery/Binding	✓	✓	✓	Long Term	✓
SlyFi: Data packets	✓	✓	✓	✓	✓

Epilogue


12000
10000
8000
6000
4000
2000
0
count
1
2
3
4
5
6
7
8
9
10
11
participant #
zip
name

Ongoing work at Intel

- Conducting a study of a day in the life of a wireless user
- Goals:
  - To understand user behavior and network activity. (To Learn what information is being sent where, when, how?)
    - What does Google OR Starbucks OR DoubleClick OR Comcast OR Joe the Plumber know about me?
  - Help determine what would actually improve privacy
    - New protocols that defend our privacy
    - Technologies to help users make more informed decisions
- Looking for student researchers to help design, implement and execute the study
  - See or mail me if interested